Will You Be My Quarantine: A Computer Vision and Inertial Sensor Based Home Exercise System*

Justin Amadeus Albert*, Lin Zhou*, Pawel Glöckner, Justin Trautmann, Lisa Ihde, Justus Eilers, Mohammed Kamal, Bert Arnrich
 {justin.albert,lin.zhou}@hpi.de
 Digital Health Center, Hasso Plattner Institute, University of Potsdam
 Potsdam, Germany

ABSTRACT

The quarantine situation inflicted by the COVID-19 pandemic has left many people around the world isolated at home. Despite the large variety of mobile device-based self exercise tools for training plans, activity recognition or repetition counts, it remains challenging for an inexperienced person to perform fitness workouts or learn a new sport with the correct movements at home. As a proof of concept, a home exercise system has been developed in this contribution. The system takes computer vision and inertial sensor data recorded for the same type of exercise as two independent inputs, and processes the data from both sources into the same representations on the levels of raw inertial measurement unit (IMU) data and 3D movement trajectories. Moreover, a Key Performance Indicator (KPI) dashboard was developed for data import and visualization. The usability of the system was investigated with an example use case where the learner equipped with IMUs performed a kick movement and was able to compare it to that from a coach in the video.

CCS CONCEPTS

• **Applied computing** → **Physical sciences and engineering**;
 • **Human-centered computing** → *Visualization design and evaluation methods.*

KEYWORDS

neural networks, computer vision, inertial measurement unit (IMU), sensor fusion, human motion tracking, 3D trajectory

ACM Reference Format:

Justin Amadeus Albert*, Lin Zhou*, Pawel Glöckner, Justin Trautmann, Lisa Ihde, Justus Eilers, Mohammed Kamal, Bert Arnrich. 2020. Will You Be My Quarantine: A Computer Vision and Inertial Sensor Based Home Exercise System. In *PervasiveHealth '20, May 18–20, 2020, Atlanta, GA, USA*. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3421937.3421946>

*Both authors contributed equally to this research.

This work is licensed under a Creative Commons Attribution International 4.0 License.
PervasiveHealth 2020, May 18–20, 2020, Atlanta, GA, USA
 © 2020 Copyright held by the owner/author(s).
 ACM ISBN 978-1-4503-7532-0/20/05.
<https://doi.org/10.1145/3421937.3421946>

1 INTRODUCTION

Due to the quarantine situation caused by the COVID-19 pandemic, many people around the world are isolated at home and unable to participate in face-to-face fitness sessions or learn a new sport with the guidance of a professional trainer. Mobile device-based home exercise tools, such as video tutorials, activity recognition and repetition counter [2], have been intensively investigated in the past years. When users perform exercises at home based on an instructional video, detailed kinematic information is needed for feedback on the accuracy of the movement compared to the tutorial. By using modern computer vision and machine learning algorithms, such as OpenPose [1] or VNect [3], skeletons can be recognized in 2D or 3D images and videos, and human motion information can be extracted.

Inertial measurement unit (IMU) is another commonly used tool for human activity tracking. An IMU typically consists of an accelerometer (measuring acceleration) and a gyroscope (measuring angular velocity). With IMUs attached to specific parts of the user's body, kinematic information such as rotation angles and movement trajectory, can be extracted using sensor fusion. Although numerous solutions exist for exercise evaluation using either computer vision or IMUs, little has been researched on comparison between both data sources [4]. In reality, movements of an exercise performed by a professional could be obtained with computer vision, while the learner only has IMUs to record their own movements, or vice versa. Since both tools provide information about a person's movement, it is possible for the learner to compare their movements to the template. However, the challenge lies between the different representations of the data. While computer vision methods continuously provide joint position data as a time series of three-dimensional joint coordinates, IMU sensors provide acceleration and gyroscope data in the local frame. To solve this problem, a method that compares data from both sources in the same format is needed.

In this paper we present a system which takes both computer vision and IMU based data as input sources, and processes the data into the same representation to compare motion. We used Kinematic data obtained from videos to predict synthetic acceleration data. In addition, we estimated the 3D trajectory of IMU sensors based on an improved algorithm tailored for repetitive motion. Both data sources can be used to gain insights in motion performance. Therefore we propose a set of Key Performance Indicators (KPI) to evaluate the quality of the performed exercises. To demonstrate the feasibility of the system, we recorded an example dataset consisting of two exercises (walking on a treadmill and kicking) with

both computer vision and IMUs, and evaluated the performance of the subjects using the proposed KPIs. The evaluation results revealed distinct differences between movements of the learner and the coach, and the KPIs.

2 METHODS

2.1 System Architecture

The overall architecture of the exercise evaluation system is shown in Figure 1. The first step of the system is to extract 2D/3D position data from a given input video to obtain information about the trajectory of the joints. Furthermore, a neural network is used to predict synthetic IMU data from 2D poses obtained from the input video. When the user repeats the exercises at home with IMUs attached, the 3D trajectory is calculated from the IMU data. By defining KPIs for each of the two representations, detailed insights into the performance of the executed exercise can be provided.

Figure 1: System overview.

2.2 Generating IMU Data from Poses

The architecture of Rey et al. [5] was used to generate synthetic IMU data based on 2D poses. In their work, 2D poses were obtained using the OpenPose pose detector, followed by a deep convolutional neural network used to predict IMU data. The original regression network first builds a shared pose representation from the sequence of 2D poses using a Temporal Convolutional Network (TCN), where then independent layers encode this representation for each sensor to obtain the final IMU data. In our architecture, 2D convolutional layers that convolved over multiple joints and time dimensions have been replaced by 1D convolutions for each sensor separately. In the original work, the magnitude of the acceleration was predicted, whereas in our case the acceleration of each individual axis was predicted, with comparable accuracy as in the original study. Figure 2 shows an example of the predicted IMU acceleration based on the modified network architecture.

2.3 Generating 3D Movement Trajectories from IMU Data

The generation of 3D movement trajectories from IMU data was based on the gait analysis algorithm developed by Tunca et al. [10].

Figure 2: Example of predicted 3-axis acceleration from given 2D human poses identified in video sequences.

An intrinsic challenge with IMU data analysis is the drift problem. To obtain information on orientation and translation (displacement), the angular velocity and acceleration recorded by the IMU would be integrated one and two times, respectively. Small errors in the raw IMU data would be magnified during the integration process and lead to large drifts in the estimations. In our algorithm, the zero velocity update (ZUPT) method was used to cope with the drift problem, where the time periods when the foot is not moving (stance phases) were identified based on the gyroscope magnitude, and thus the errors from the raw IMU data could be characterized [9]. An error-state Kalman filter tracks and corrects the sensor errors, and produces estimates of the 3D movement trajectories of the feet. In addition to the original algorithm which was intended for IMU gait analysis, our implementation of the algorithm also takes the type of exercise into consideration. When the subject is doing exercise at a stationary position (e.g., practising a Taekwondo kick from the same standing position), the starting point in the estimated trajectory for each repetition is reset to the origin, thus creating a clear visualisation of the variations in all repetitions. Figure 3 shows an example of the estimated trajectories.

2.4 KPI Dashboard for Quality Assessment

A set of Key Performance Indicators (KPIs) were developed in order to quantify the movements captured by both the computer vision and the IMU systems. KPIs derived from the raw IMU data of each repetition include duration, mean and standard deviation of the acceleration, and average speed. KPIs derived from the 3D movement trajectory of each exercise repetition include vertical and horizontal movement distances and the 3D path length.

In addition, we defined a new metric, 3D path efficiency, as shown in the equation 1. It is the ratio between the Euclidean

Figure 3: 3D trajectories of multiple kicks with aligned starting positions.

distance d between the origin and the maximum reach point of the trajectory and the actual foot trajectory s . When the 3D trajectory is a straight line (i.e. very efficient), the 3D path efficiency is 100%. In case of stationary repetitive movements (e.g. kicking), the ratio is multiplied by two to cancel out the doubled distance.

$$\text{3D Path Efficiency} = \frac{2d}{s} \quad (1)$$

Table 1 summarizes the KPIs developed in this study. In addition, a dashboard was developed to visualize the motions captured by the computer vision and the IMU systems, and to compare the KPIs from different movement recordings. In the dashboard interface, the user is able to load computer vision and IMU data into the analysis pipeline, select an arbitrary part of the data to be visualized as raw IMU data and movement trajectories. The corresponding KPIs for the selection are displayed next to the data visualization.

Table 1: KPIs derived from raw IMU data and 3D trajectory.

Data Source	KPI
Raw IMU Data	Mean Acceleration (m/s^2)
	Standard Deviation Acceleration (m/s^2)
	Mean Speed (m/s)
	Mean Repetition Duration (s)
3D Trajectory	Vertical Movement Distance (m)
	Horizontal Movement Distance (m)
	3D Path Length (m)
	3D Path Efficiency (%)

3 EVALUATION

3.1 Dataset

An example dataset was recorded to evaluate the movement comparison system developed in this study. The dataset contained two different exercises performed by three participants. The first exercise was treadmill walking for one minute at three different velocities. The second activity was repetitive kicking at three different

heights. The participants were young and healthy male adults with an average age of 22.7 ± 1.2 years, body height of 184 ± 10.1 cm and body mass of 74 ± 6.9 kg. In this study, a Kinect v2 camera (Microsoft, USA) was placed in front of the subject to capture RGB images as well as 2D and 3D joint positions from an integrated skeletal tracking algorithm [8]. Seven Physilog[®] 5 IMUs (Gait Up, Switzerland) were attached to the participants (two on top of the feet, two below the knees, two on the thighs, and one on the sacrum). The sampling rate of the IMUs were 128 Hz, accelerometer and gyroscope ranges were $\pm 8g$ and $\pm 2000 dps$, respectively. The Kinect data was recorded at 30 Hz and up-sampled to 128 Hz for analysis. Figure 4 shows one subject performing walking on a treadmill and performing a kick while standing at a fixed position.

Figure 4: Experimental setup. Left: treadmill walking. Right: kicking from a fixed position.

3.2 Movement Comparison with the KPI Dashboard

To simulate the situation where a learner equipped with IMUs is practicing movements taught in a video tutorial, we selected Kinect data from a subject who has experience in martial arts (as the coach) performing high kicks and IMU data from another inexperienced subject (as the learner) performing high kicks. The KPI dashboard was used for data loading and exploration. Figure 5 is a screenshot of the trajectory section in the dashboard with one example kick from each subject. The learner had a smaller mean speed and a larger movement duration compared to the coach. This was in line with our expectations, as the learner moved slower than the coach for the analyzed kick. The KPIs derived from the 3D trajectories revealed that the vertical and horizontal distances as well as the 3D path length of the kick from the learner were smaller than those from the coach. Moreover, the 3D path efficiency, defined by the ratio between the Euclidean distance of two ends of the kick and the 3D path length, was smaller for the learner than for the coach. A closer analysis on the shapes of the trajectories revealed that the trajectory from the coach was more directed and contained less unnecessary turns compared to the one from the learner. With the 3D trajectory visualization in the dashboard, it was easy for the user to look at the movement from both the coach and the learner in

Figure 5: Screenshot of the trajectory section on the KPI dashboard. 3D trajectories and derived KPIs from the coach with computer vision data (in blue) and the learner with IMU data (in orange) could be clearly visualized.

arbitrary angles, and identify the exact differences between the two movements. However, in this particular case, the exact trajectory from the coach should be studied with caution, as the low sampling rate from the Kinect system led to a generally less smooth trajectory with occasionally unexpected sharp edges.

4 DISCUSSION

This study presents a computer vision and inertial sensor based home exercise system that enables the user to compare exercise movements from different data sources. The implementation of the generated synthetic IMU data from poses was able to predict acceleration in 3-axis, and achieved comparable accuracy to the acceleration magnitudes reported in the original study. The KPI dashboard developed in this study works as a user-friendly interface for data loading, exploration and movement analysis. With the set of KPIs alongside both raw IMU data and 3D trajectory, differences between movements of the learner and the coach are clearly quantified, and visualized in an intuitive way. In this study, we evaluated the system using an example kicking movement. However, additional exercises can be imported and analyzed using the dashboard. Depending on the types of exercises, additional KPIs from different modalities can also be integrated into the existing framework. For example, joint angles and ground reaction force are commonly used modalities for quantifying human motion, and can be derived from both computer vision and IMU data [6, 7]. It is worth emphasizing that the current study is a proof of concept for the proposed home exercise system. The testing dataset was recorded with only a small number of subjects and a limited set of exercise movements. For the purpose of demonstrating the functionalities of the system, only one use case was analyzed. To confirm the general usability of the system, further analysis on a larger dataset with more subjects and multiple use cases is needed.

5 ACKNOWLEDGMENTS

This research has been partly funded by the Federal Ministry of Education and Research of Germany in the framework of KI-LAB-ITSE (project number 01IS19066), and the HPI Research School on Data Science and Engineering.

REFERENCES

- [1] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh. 2019. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2019).
- [2] Preetham Ganesh, Reza Etemadi Idgahi, Chinmaya Basavanahally Venkatesh, Ashwin Ramesh Babu, and Maria Kyrarini. 2020. Personalized System for Human Gym Activity Recognition Using an RGB Camera. In *Proceedings of the 13th ACM International Conference on Pervasive Technologies Related to Assistive Environments* (Corfu, Greece) (PETRA '20). Association for Computing Machinery, New York, NY, USA, Article 21, 7 pages.
- [3] Dushyant Mehta, Srinath Sridhar, Oleksandr Sotnychenko, Helge Rhodin, Mohammad Shafiei, Hans-Peter Seidel, Weipeng Xu, Dan Casas, and Christian Theobalt. 2017. VNect: Real-time 3D Human Pose Estimation with a Single RGB Camera. In *ACM Transactions on Graphics*, Vol. 36. ACM New York, NY, USA, 1–14. Issue 4.
- [4] Bojan Milosevic, Alberto Leardini, and Elisabetta Farella. 2020. Kinect and wearable inertial sensors for motor rehabilitation programs at home: State of the art and an experimental comparison. *BioMedical Engineering Online* 19, 1 (2020), 1–26.
- [5] Vitor Fortes Rey, Peter Hevesi, Onorina Kovalenko, and Paul Lukowicz. 2019. Let There Be IMU Data: Generating Training Data for Wearable, Motion Sensor Based Activity Recognition from Monocular RGB Videos. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers* (London, United Kingdom) (UbiComp/ISWC '19 Adjunct). Association for Computing Machinery, New York, NY, USA, 699–708.
- [6] Thomas Seel, Jörg Raisch, and Thomas Schauer. 2014. IMU-based Joint Angle Measurement for Gait Analysis. *Sensors (Switzerland)* 14, 4 (2014), 6891–6909.
- [7] E. Shahabpoor and A. Pavic. 2018. Estimation of vertical walking ground reaction force in real-life environments using single IMU sensor. *Journal of Biomechanics* 79 (2018), 181–190.
- [8] Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Toby Sharp, Mark Finocchio, Richard Moore, Alex Kipman, and Andrew Blake. 2011. Real-time human pose recognition in parts from single depth images. *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition* (June 2011), 1297–1304.
- [9] Isaac Skog, Peter Händel, John Olof Nilsson, and Jouni Rantakokko. 2010. Zero-velocity detection-An algorithm evaluation. *IEEE Transactions on Biomedical Engineering* 57, 11 (2010), 2657–2666.
- [10] Can Tunca, Nezihe Pehlivan, Nağme Ak, Bert Arnrich, Gülüstü Salur, and Cem Ersoy. 2017. Inertial sensor-based robust gait analysis in non-hospital settings for neurological disorders. *Sensors (Switzerland)* 17, 4 (2017), 1–29.