Dear Reviewers,

At the beginning we would like to thank reviewers for their effort and their comments. According to them we have improved the article, however we would like to answer them directly in this letter as well.

**How do you define your IMU frame with respect to the limb measured. Where is your Kinect reference frame? How do you define the relative orientation between the two different systems?**

A reference frame, which defines the startup pose, is the frame where the user stands in the T-pose at the beginning of each exercise. Such pose has been chosen as easy to recognize by Kinect controller (no occlusions, all joints visible and tracked), as well as neutral for IMU (no bones rotations in such pose).

Both systems, Kinect and IMU, use different coordination space with axes defined in a different way. Kinect and IMU coordination spaces have been presented in the Figure 1 (Figure 12 from the manuscript ).

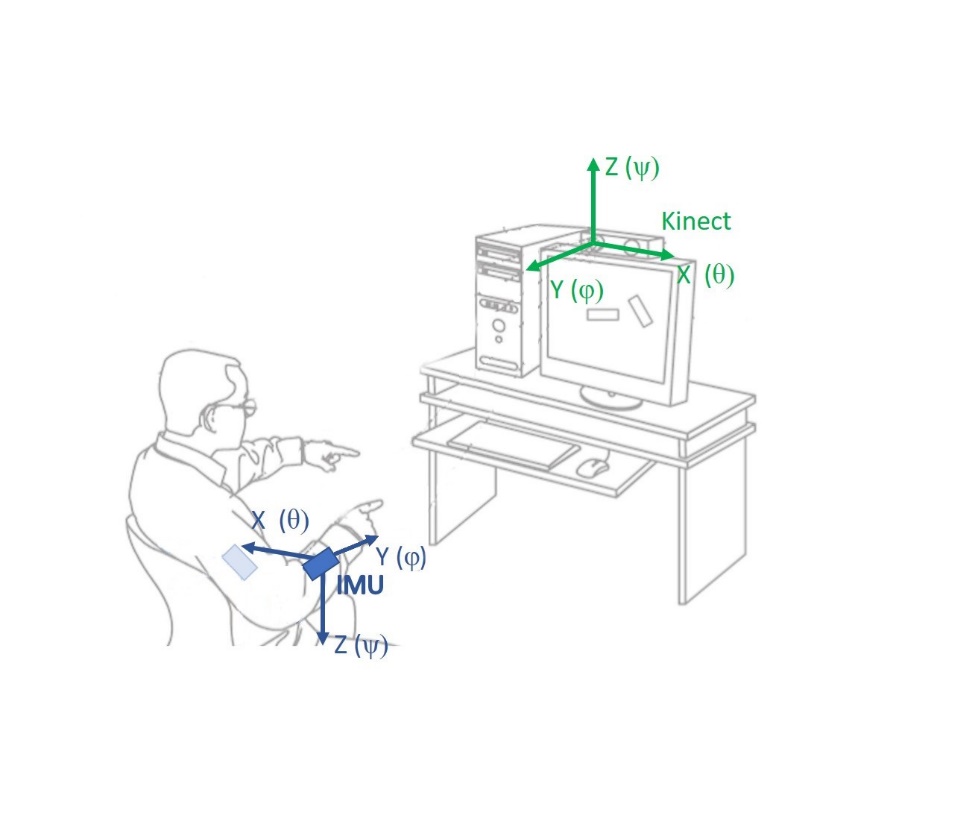


Figure 1 Kinect and IMU coordination spaces used in proposed methods.

The coordination space for fused data has been defined in direct relation to the Kinect’s space, so IMU’s measurements must be transformed into Kinect coordination space as to unify the values and avoid ambiguity. As IMU coordination system is anchored to the gravity vector, and retains its orientation regardless of device rotations, to achieve unification, a 3D orientation estimated basing on accelerometer and gyroscope measurements:

(

needs to be rotated first by 180° over the Y axis () and then reflected through the XZ surface (scaled Y axis measurement by factor -1, ). The descripted transformation is presented with the following equation:

=

After the measurements unification, they can be fused together unambiguously.

To estimate the 3D position of each joint, the hierarchical model of the limb has been defined - hand serves as an example with the root in the shoulder joint. The model has been presented in the Figure 3 (Figure 9 from the manuscript ).

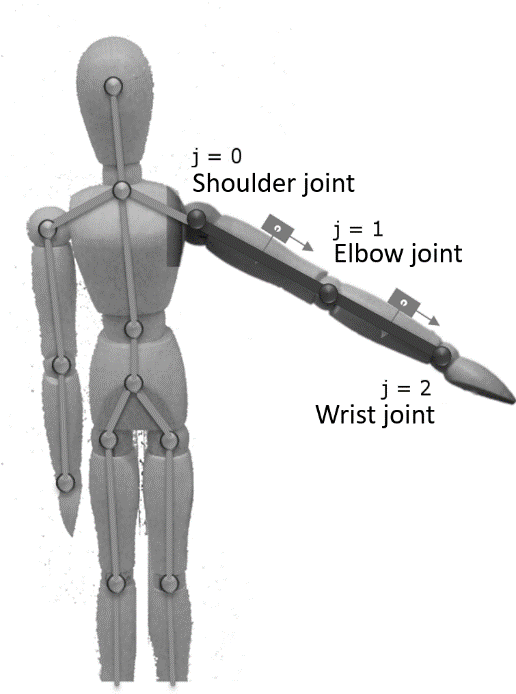


Figure 3. Hierarchical upper limp model definition with the IMU devices locations

The 3D position of the root joint is taken directly from the Kinect coordination system. Then the elbow joint position is calculated basing on the fused arm bone estimation and arm bone length (estimated once during calibration). Consecutive steps are performed for subsequent joints of skeleton kinematic chain - the position of the wrist joint is calculated analogically to the elbow’s position.

Article lines: 263 - 276

**In the Euler angle filtering, in 2.3.5. you fuse two Euler angles. How can Kinect provide Euler angle, and what is the common frame used to make a meaningful data combination?**

We assume that the question concerns directly to the chapter 2.3.6. In available skeleton data stream, the bones’ orientations, represented in the form of quaternions, are inherent data types provided by the Microsoft Kinect controller. These values are real-time calculated basing on the joints’ estimated positions and are updated with the joints’ positions update. If we assume that quaternion () is defined as where is the real part of the quaternion number, then conversion from the quaternion to Euler angles is defined with a following equation:

Analogical procedure is required to convert the Madgwick’s filter results, which originally describe 3D space orientation of IMU device in the quaternion form, into Euler angles representation. Resulting orientations (described with Euler angles), transformed into a common frame of Kinect controller device can be fused according to the elaborated, context-based, weighted data fusion method.

**In line 320, the fusion parameters are defined as [0.98,0.05,0.65], does this parameter apply for other types of motions?**

The limbs motions have been chosen as the most challenging for exploited motion controllers. Selected upper limbs moves emphasize devices limitations. For example, during the elbow’s flexion to the front (motion numbered in the article as #2), some joints become hidden, so Kinect loses its tracking ability. Moreover, during such move IMU is not able to detect it correctly, as the limb rotation is around the gravity vector. The proposed method coefficients, elaborated basing on self-experiments and measurement devices characteristics, not only handled such scenario, but was proved to be optimal for selected parameters (i.e. temperature, joint depth distance, joint reliability). Bearing in mind that proposed exercises represent motion trajectories that can be considered as difficult to track, by the used class of measurement devices, it is justified to claim that any other motion would be easier. For any other measurement devices from the selected classes of precision/reliability (industrial devices for IMU [Alexiev2013] and structured light map based depth cameras for home entertainment), as the one we used in our research, proposed coefficients should be universal and applicable for other types of human limbs motions.

[Alexiev2013] - K. Alexiev and I. Nikolova, “An algorithm for error reducing in IMU,” *2013 IEEE Int. Symp. Innov. Intell. Syst. Appl. IEEE INISTA 2013*, no. 1, 2013.

The description of experiments and their results, that allowed to define mentioned values, has been extended in the manuscript in lines 369-380.

**Please comment the difference your research from the paper that is "Upper limb motion tracking with the integration of IMU and Kinect", Yushuang Tian et all, Nurocomputing 159, 207-218. And I need to know enhancements or novelties of your paper w.r.t Yoshuang Tian's paper. Because your references don't have Yoshuang Tian's paper.**

Indeed, Mr. Tian’s article addresses the similar problem of the upper limb motion tracking with IMU and Kinect devices. However, there are several aspects that differ our method and allow to consider our solution as a novel approach.

From the method design point of view, we would like to emphasize 4 main differences.

1. The first difference is related to the accelerometer and gyroscope data fusion. The Tian’s method uses Unscented Kalman filter to calculate the IMU device orientation, while our method is based on the Madgwick’s filter, which appeared to be reliable.
2. Unscented Kalman filter is used by Tian for fusion of Kinect’s data stream with data calculated from IMU device. To perform such fusion we propose an approach based on the data reliability level and importance of each value, based on the context of the motion.
3. The controllers data stream has a different representation, what constitutes one of the main contributions of our method. Tian proposed to fuse joints positions estimated by Kinect with joints positions estimated by the IMU sensor. In opposition to that approach, we proposed more beneficial bones orientations data fusion. Such approach allows to fuse less processed data (bones orientation) in comparison to the position based approach, as it eliminates the additional source of error (bone length estimation and joints positions fluctuations) in data fusion. Moreover, our approach is based on the precalculated, at the calibration stage, hierarchical human skeleton model, which includes information about necessary lengths of each segment.
4. The last difference that we would like to emphasize is related to the improvement of the measurement data quality and its reliability. Tian’s method focuses mostly on geometrical constraints of the performed motion. It means that authors limit the influence of the joints positions estimations that are impossible to achieve during the Kinect-IMU data fusion. In our approach, we focus on the characteristics of measurement devices and their imperfections, as well as on the methods of the contextual compensation of such errors.
5. The last contribution is that the characteristics of exploited devices were not described in the literature, in the complex way. This field may be also considered as the novel input to the overall discipline’s knowledge.

Additionally, the types of the exploited measurement devices differentiate these two methods. InertiaCube3, used by Tian, is built from an accelerometer, a gyroscope and a magnetometer, what means that this device should be considered as MARG device rather than IMU (an accelerometer and a gyroscope only), which was used in our method (MPU-6050). This fact is important while considering the complexity of the performed tests. In the Tian’s experiments, the only possible measurement problem was related to human skeleton joints occlusion, which is one of the characteristic imperfections of Kinect device. In our approach, we additionally considered the lack of the rotation around the gravity vector. This implies a different approach to the data fusion, due to the fact of the incompleteness of the information from both sources.

Tian in his article hasn’t published the final accuracy of his method. From the published position estimation diagrams it can be estimated that the average positioning error is about 3cm, while our method has about 2.5cm average positioning error for the elbow joint and 2.8cm for the wrist joint. The only numerical estimation value, presented in Tian’s paper, relates to the elbow adduction angle estimation deviation, which had its maximum of about 20 degrees. In our method, the average elbow abduction angle estimation error varies from about 4 degrees up to 15 degrees, depending on the hand motion direction and its measurement difficulties, what makes our approach more precise in this aspect.

Discussion about Tian’s article has been also included in current revision of the article. Article lines: 80-87

We believe that the mentioned differences are sufficient to prove that, even though both methods address the similar problem of the upper limp motion tracking, our method presents different and more precise approach to IMU and Kinect data fusion.

Current revision of the article has been also edited in crucial parts that, we believe, help to follow proposed method. Both new or edited parts have been highlighted in the article.

If we can help and clarify more aspects we will be more than happy.

Regards

Authors