

Hybrid Method of Human Limb Joints Positioning—Hand Movement Case Study

Grzegorz Glonek and Adam Wojciechowski

Abstract Precise and unambiguous limbs motion tracking is one of the key aspects laying behind natural human-machine communication. The paper presents a novel approach to depth sensor (Microsoft Kinect) and inertial measurement units (IMU) data fusion, providing more precise and stable hand joints tracking. The new method substitutes, mainly described in literature, sensors-derived joints position fusion with sensors-derived bones orientations fusion and subsequent joints positions estimation. Obtained joints positioning precision became even 25 % better than in other solutions. The paper comprises also the method evaluation results. It was verified both against professional motion tracking VICON system and Kalkbrenner method [6], the most relevant to the presented solution.

Keywords Kinect · IMU · Fusion · Biomechanics · Joints positioning

1 Introduction

Limb motion tracking, understood as an unambiguous and delay-minimizing process of limb's joints 3D space position estimation, is a valid problem invaluable for current researches on Natural User Interface design. It is used nowadays in several areas such as entertainment (games and movies animations), interaction with scene objects in augmented reality systems or motor skills rehabilitation. This last area, supported by the computer system, constitutes a part of a wider subject named tele-rehabilitation.

For several years, the only possibility to obtain the limb joints tracking desired accuracy was to exploit professional motion capture systems i.e. VICON or Optitrack. However, since couple of years, there appear broadly available (and cheaper) devices (i.e. game controllers) that allow to track selected aspects of human motion at user's home. On the other hand, nonprofessional solutions reveal several imprecisions and

G. Glonek (✉) · A. Wojciechowski
Institute of Computer Science, Lodz University of Technology, Lodz, Poland
e-mail: grzegorz@glonek.net.pl

constraints that might be compensated by an appropriate controllers derived data fusion.

In the paper two types of such devices were taken into consideration: Microsoft Kinect 360—a RGB–D camera that is able to track whole human body and inertial measurement units (IMU) consisting of accelerometer and gyroscope sensors that are able to measure linear acceleration and angular speed. As the devices' recording frequencies are limited (30Hz for Kinect and 70Hz for IMU) and the context of rehabilitation and hand gesture controlled object manipulation is considered, the hand tracking accuracy is superior to the speed of hand movement.

Though several authors [2–4] have proved that Kinect and IMU data fusion assures limbs joints positions tracking accuracy of about 2.5–3 cm, presented method achieves better results: 2–2.5 cm.

2 State of Art

Considered sensors have several measurement characteristics that should be taken into consideration during the fusion. Microsoft Kinect controller loses its tracking ability due to body parts occlusion [1]. Moreover, while tracking, lost joints may affect the tracking accuracy of those which are fully visible to the sensor's camera. The rotation of user's body might be an example of such scenario. Basing on author's experiments, if the user rotates more than 50° (angle between the user and camera view directions), occluded shoulder joints (and almost half of the body) will be invisible to the device and measurements of visible parts will be unstable.

Another important characteristic is that joints positions measurement accuracy change with the distance between the human and the device [7]. Figure 1 shows how the estimated accuracy changes with a distance. As the most important IMU flaws, the gyroscope measurement drift and the temperature related bias in accelerometer

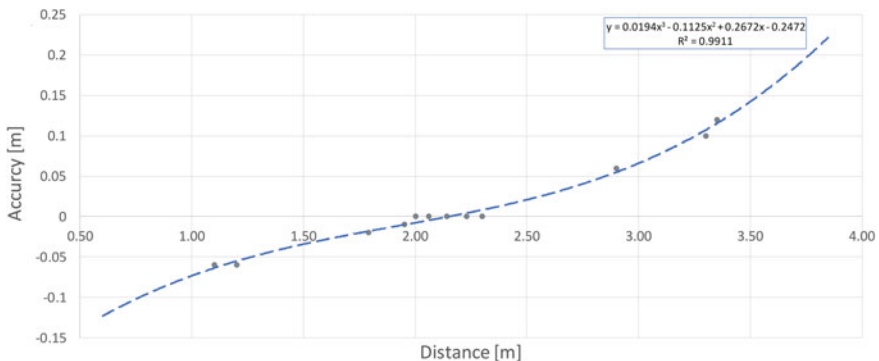


Fig. 1 Position measurement accuracy in Z-axis to distance from the Kinect

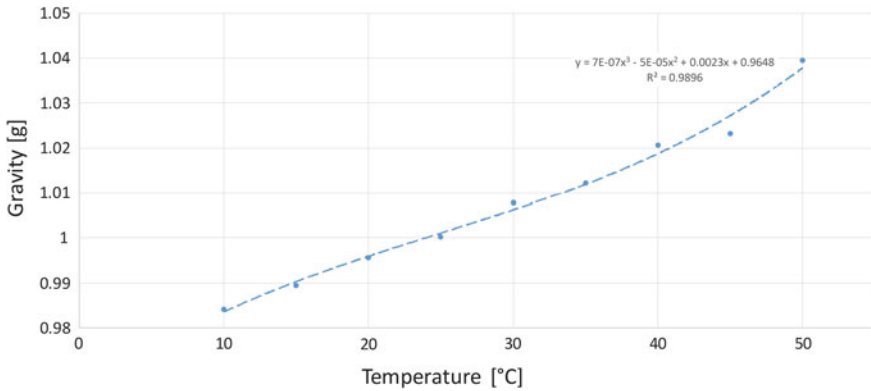


Fig. 2 Gravity measurement in temperature range 10–50°C

measurements may be indicated [11, 17]. The influence of the temperature on the bias is presented in Fig. 2.

The studied paper [9] estimates Microsoft Kinect general posture estimation variability in range 5–10 cm. Moreover, the author pointed out that the length of tracked bones vary between measurement frames in range 2–5 cm.

Considering selected controllers, several hybrid data fusion approaches, improving positioning accuracy, can be found in the literature. Authors have elaborated different approaches to sensors’ data fusion, which characterize various levels of measurements reliability. The first group of approaches can be classified as methods which use Kinect measurements as a reference system and partially relay on its measurements.

Bo et al. [2] described the joint angle (angle between joint adjacent bones) estimation method exploiting 5 degrees of freedom (DOF) IMU and Kinect. In the initial stage the method interprets the gyroscope and the accelerometer data separately and their fusion by the linear Kalman filter (KF) is subsequently performed. The same angle estimated with Kinect data is used to initially calibrate and then temporarily correct the bias of the accelerometer estimation. There were no numerical results provided in the paper, however the presented charts suggest considerable improvement of data fusion results.

A different approach was presented in a paper published by Destelle et al. [3]. Authors decided to use a set of 6DOF IMUs supported by a magnetometer, where each unit was stuck to one of the tracked limb bones. Basing on gathered measurements, orientation of each bone was estimated by the Madgwick’s algorithm [8] and their superposition resulted in the full skeleton model. Kinect data was used twofold. The first stage was to get the initial, reference skeleton frame to *label* data estimated from inertial units and to improve the IMU calibration. That process resulted in the hierarchical definition of bones orientations (*inertial skeleton*). The second stage of Kinect exploitation was to track the position of the central body point (torso joint) and then update the whole *inertial skeleton* relatively to this points displacement.

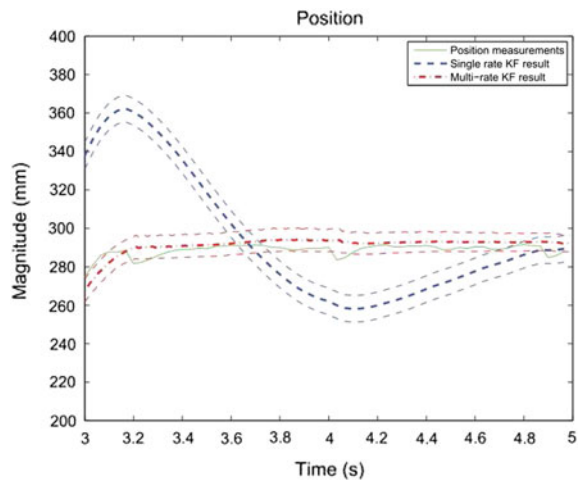
Resulting, VICON-referenced knee joints angle estimation error varied between 4° and 14° . It depended on cross correlated joints, the measurements were referenced to, as there were no joints absolute angle estimation performed.

The newest method that could be qualified to this group, is the one presented in 2015 by Tian et al. [12]. Authors included geometrical constraints of the performed motion in the estimation process, to eliminate estimations that are impossible to achieve in the real life i.e. angle between forearm and arm cannot be greater than 180° . The fusion algorithm used by authors is based on the Unscented Kalman Filter (UKF) [15]. Presented results show that the algorithm is able to work also while Kinect is outage, which was not obvious in previously described methods. To estimate method accuracy, authors compared elbow angle measurements. Authors published the information that angle measurements estimated by their fusion method deviates less than 20 degree from the expected value.

The another group of published methods is based on the assumption that both measurement devices' imperfections need to be corrected continuously by the signal fusion. In 2014 Feng and Murray-Smith [4] proposed the multi-rate Kalman filter-based fusion method of joints positions, estimated by Kinect with linear acceleration and velocity of this joint. Presented results showed that this approach stabilizes measurements around the real value much faster than single rate KF. This is visible especially when the movement starts/stops (Fig. 3). The accuracy of the presented method can be estimated around 1.5–2 cm (based on published diagrams). However, presented results refer to very short time periods (up to 5 s), so it is impossible to estimate how the method works in a long term.

A different approach was presented in the paper of Kalkbrenner et al. [6]. Authors of this publication suggested Kalman-based linear fusion of joints estimated positions retrieved from bones orientations superposed with the skeleton model (bones length)

Fig. 3 Joint position estimation with the multi-rate Kalman filter and the single rate Kalman filter [4]. *Red line*—multi-rate Kalman filter, *blue line*—single-rate Kalman filter



and Kinect measurements. Absolute joints positions estimation results were around ± 2.5 cm and seem to be the most accurate in long term experiments.

In the papers presented so far, motions, described as test movements, were performed in the *Kinect friendly* way. It means that all tracked joints were fully visible during every motion sequence and movements were done along the single axis and didn't include any occlusions.

The approach presented in 2013 by Helten et al. [5] is more advanced than methods presented so far and uses the pattern recognition to estimate the user's current pose. In the previously analyzed articles, the IMU data was always fused with Kinect skeleton data and basing on that, multiple pose features were calculated. In this article, Kinect is used as a depth camera and data from the depth stream is fused with IMU measurements. The motion recording was performed with the use of six Xsens MTx 9 DOF IMUs [18] which is a full set of professional inertial motion tracking system. The proposed method is based on the idea of the visibility model that is built from poses estimated on inertial and Kinect measurements, and then matched with predefined poses stored in the database. However, such approach seems to be useless in scenarios where you need to estimate joints positions and other limb features, as it focuses on the general pose recognition.

Our method is the most similar to the Kalkbrenner approach, but our main contribution consists in improved, weighted sensor contextual influence which results in overall better absolute joints positions estimations. The method joints absolute positioning precision of about 2–2.5 cm was verified against the ground-truth VICON system.

3 Method

A method proposed by authors, bases on the continuous linear fusion of skeleton bones orientations with respect to the current motion context. It takes into consideration controllers reliability and compensates evaluation imperfections. The proposed motion positioning method can be defined as a function f :

$$f(A, G, T, P_0^K, P_1^K, Q^K, \Delta t) \Rightarrow [p_x^F, p_y^F, p_z^F]_t, \quad (1)$$

where: A —accelerometer measurement, G —gyroscope measurement, T —temperature measurement, P_0^K, P_1^K —start and end bone joints positions measured by Kinect i.e. elbow and wrist, Q^K —bone orientation estimated by Kinect, t —current time frame, Δt —elapsed time between previous and current frame.

Orientations are contextually presented in two forms: quaternions and Euler angles, and they are transformed between these forms with respect to North and South Pole singularities. In the method authors exploited limbs joints positions ($P^K = [p_x^K, p_y^K, p_z^K]$) and bones orientations ($Q^K = [q_w^K, q_x^K, q_y^K, q_z^K]$) supplied by the Kinect device as well as accelerometer ($A = [a_x, a_y, a_z]$), gyroscope

($G = [g_x, g_y, g_z]$) and temperature (T) measurements from each IMU. Kinect joints positions and IMU based marker locations on tracked limbs are presented in Fig. 6.

In the proposed method, data gathered from measurement devices, are denoised in the first step and then used to calculate bones orientations. Then, orientations calculated from IMU devices and measured by Kinect are fused together and in the last step, bones length model is added to estimate absolute joints positions. The general overview of the orientation-based fusion process is presented in Fig. 4.

The data processing is performed in two parallel threads. The first one performs computations on the IMU data to estimate limbs orientations (quaternion) and the second retrieves Kinect skeleton bones orientations (quaternion). Their consequent, contextually-weighted and time-correlated, superposition results in fused bones quaternions values which, assuming skeleton model, can be transformed into estimated joints absolute positions. At the beginning of the first thread, the IMU accelerometer bias was corrected with the (2), due to the device operating temperature destructive influence on its measurements.

$$A' = \frac{A}{1 + \beta(T - T_0)}, \tag{2}$$

where:

- A' —corrected accelerometer measurement,
- A —accelerometer measurement,
- T —temperature measurement,
- T_0 —device reference operating temperature. For used device $T_0 = 25^\circ\text{C}$,
- β —correction factor. For used device $\beta = 0.0011$.

The value of β correction factor is the result of exploited IMU gravity regression analysis as a function of the device operating temperature (Fig. 2). Next, the corrected accelerometer data and gyroscope measurements were used to calculate quaternion of adjacent bone orientation with the Madgwick's filter [8]. The estimated orientation was then extrapolated to eliminate the observed delay. The linear extrapolation algorithm was used in this step (3).

$$[\Phi, \Theta, \Psi]' = [\Phi, \Theta, \Psi]_t + \gamma([\Phi, \Theta, \Psi]_t - [\Phi, \Theta, \Psi]_{t-1}), \tag{3}$$

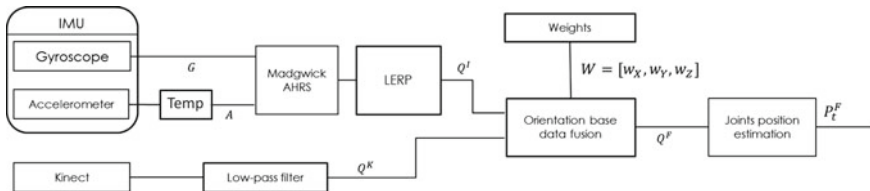


Fig. 4 Orientation based fusion method

where: $[\Phi, \Theta, \Psi,]'$ —corrected orientation in the form of Euler angles; $[\Phi, \Theta, \Psi,]$ —orientation in the form of Euler angles; γ —extrapolation factor. For used device $\gamma = 0.5$.

In the second thread, Kinect data needed to be denoised with no significant delay in measurements. It was done by first-order exponential low-pass filter defined by Eq. 4. Both joint positions and orientations have been filtered in this step.

$$y_t = \alpha x_t + (1 - \alpha)y_{t-1}, \tag{4}$$

where: y —filtered data; x —noised data; α —filtration factor. $\alpha = 0.065$.

The α factor value was estimated as a result of the analysis of the average Kinect positioning error during hand motion sequence (Figs. 5 and 6).

Both devices work in different coordination spaces (Fig. 7) and they need to be transformed into the common space before their data can be fused. As the majority of data is gathered from Kinect, its coordination space was chosen as the main one. That minimized additional computations that need to be done.

In the next step, the controllers' quaternions fusion was performed. It started with the assessment of Kinect measurements reliability. The user orientation to the camera and joints positions measurement noise level were taken into consideration. The noise level is measured by the high-pass filter in the form of Eq. (5). Sample results for keeping hands with no motion and for with lost tracking are presented in Fig. 8.

$$n_t = \delta n_{t-1} + \delta(P_t - P_{t-1}) \tag{5}$$

[16] where:

- n —noise level,
- P —joint position,
- δ —filtration factor $\delta = 0.01$.

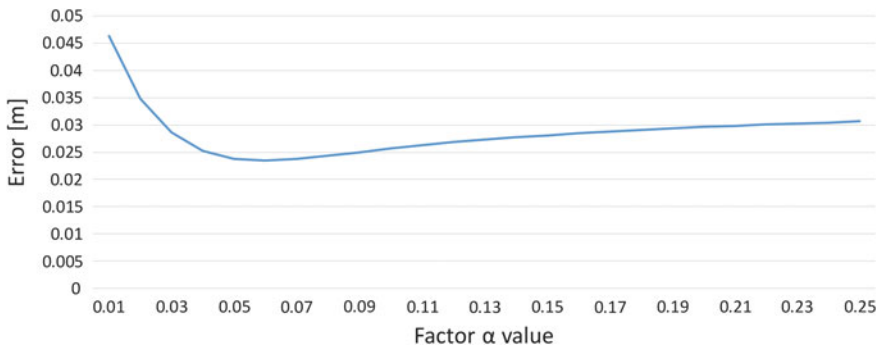


Fig. 5 Kinect measurement accuracy to low-pass filter factor α

Fig. 6 Kinect skeleton joints positions and IMU location

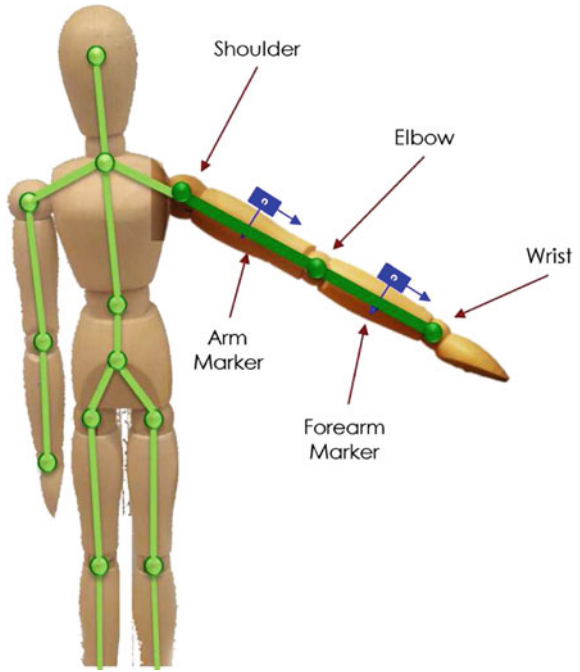
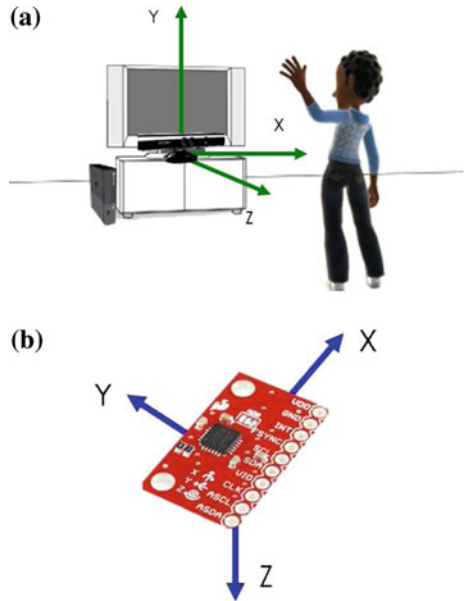


Fig. 7 Devices coordination spaces. **a** Kinect. **b** IMU



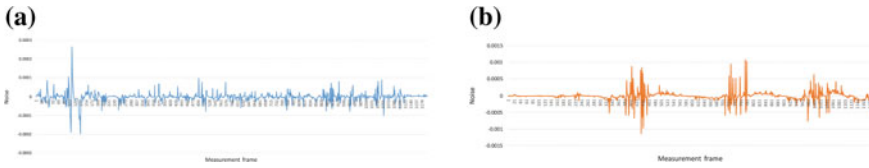


Fig. 8 High-pass filter (Eq.5) results for joint position while joint is not occluded in any time (a) and partially occluded (b). **a** Not occluded joint. **b** Occluded joint

If the user is rotated to Kinect camera more than 50° (the angle between the line of user’s shoulders and the camera surface) or the noise level is too high ($|n| > 0.0004$ based on performed experiments) then Kinect measurements are classified as unreliable and are replaced with the difference between current and previous IMU-based orientation estimations. The orientations fusion is defined as the complementary process where rotations around each axis are joined with different weights. This approach was motivated by the fact that the controllers have different measurement abilities and accuracy along each axis.

If current Kinect’s data is classified as reliable, the fusion is expressed by the following Eq. (6).

$$\begin{bmatrix} \Phi^F \\ \Theta^F \\ \Psi^F \end{bmatrix}_t = \begin{bmatrix} w_x & 0 & 0 \\ 0 & w_y & 0 \\ 0 & 0 & w_z \end{bmatrix} \begin{bmatrix} \Phi^I \\ \Theta^I \\ \Psi^I \end{bmatrix}_t + \begin{bmatrix} 1 - w_x & 0 & 0 \\ 0 & 1 - w_y & 0 \\ 0 & 0 & 1 - w_z \end{bmatrix} \begin{bmatrix} \Phi^K \\ \Theta^K \\ \Psi^K \end{bmatrix}_t \quad (6)$$

where: $[\Phi^X, \Theta^X, \Psi^X]^T$ —Euler form-based orientation; X denotes: F —fused, I —IMU, K —Kinect, and w_x, w_y, w_z : weights defining fusion factor of each axis rotation. Defined as $[0.98, 0.05, 0.65]$ respectively.

Weights used in Eq. 6 describe the level of importance of IMU measurements and need to be < 1 . The higher value used, the more important measurement was. As Kinect is not able to measure rotation along ‘x’ axis (Roll), weight close to 1 has been used there. In case of usage of inertial sensors without magnetic sensor support, rotation around gravity vector (‘y’ axis—Yaw) contains uncompensated time-related drift, so, in this case, IMU measurement was discriminated. Third axis rotation was measured by both devices, however IMU had slightly better accuracy than Kinect which was reflected in weight > 0.5 .

As both devices, Kinect and IMU, work with different sampling frequency, decimation technique has been used to pick the measurements from the closest time frames.

In case of Kinect data unreliability, the fusion was made between previously fused value and the IMU-based orientation update—between the previous and the current measurements. The fusion formula is defined as follows (7):

$$\begin{bmatrix} \Phi^F \\ \Theta^F \\ \Psi^F \end{bmatrix}_t = \begin{bmatrix} \Phi^F \\ \Theta^F \\ \Psi^F \end{bmatrix}_{t-1} + \begin{bmatrix} w_x & 0 & 0 \\ 0 & w_y & 0 \\ 0 & 0 & w_z \end{bmatrix} \left(\begin{bmatrix} \Phi^I \\ \Theta^I \\ \Psi^I \end{bmatrix}_t - \begin{bmatrix} \Phi^I \\ \Theta^I \\ \Psi^I \end{bmatrix}_{t-1} \right) \quad (7)$$

In this case w_x value remains the same and it equals 0.98 and w_y and w_z get low over the time according to the following function (Eq. 8):

$$w = \left(1 - \frac{\Delta t_n}{10} \right) \cdot 0.65, \quad (8)$$

where w —current weight value, Δt_n —amount of time in seconds when Kinect stays unreliable.

When the controllers-fused orientation is estimated it must be recalculated to the quaternion form. Then, basing on the known bone length, the position of the desired joint might be calculated.

4 Results

In order to verify the elaborated method (orientation-based joints position estimation) precision several experiments were performed. They were conducted with the VICON motion capture system as a source of a ground-truth reference data. Five users were monitored with Kinect controller, two IMUs attached to arm and forearm bones (Fig. 6) and passive marker-based VICON system, simultaneously. Markers were attached to the hand according to the schema presented in Fig. 9.

The PC used to record Kinect and IMU data was a 2.5 GHz Intel Core i7-4710HQ CPU base computer with 8GB of RAM and SSD drive. The exploited Kinect device was a dedicated Xbox 360 console controller. The software was implemented on .Net Framework 4.5 with Kinect SDK v. 1.8. IMUs—these were MPU6050 devices set up with scale ranges: ± 4 g for accelerometer and $\pm 500^\circ/s$ for gyroscope. Inertial

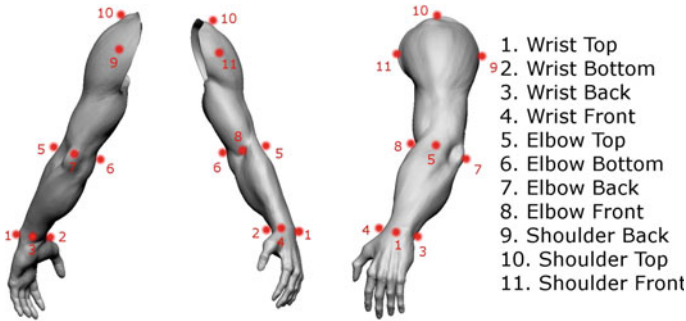


Fig. 9 Used VICON Arm marker model (Pai [10])

devices worked as a part of the self-made measurement device, built on the Arduino Due platform. The data transmission between IMUs and the PC was done through Bluetooth v. 2.0.

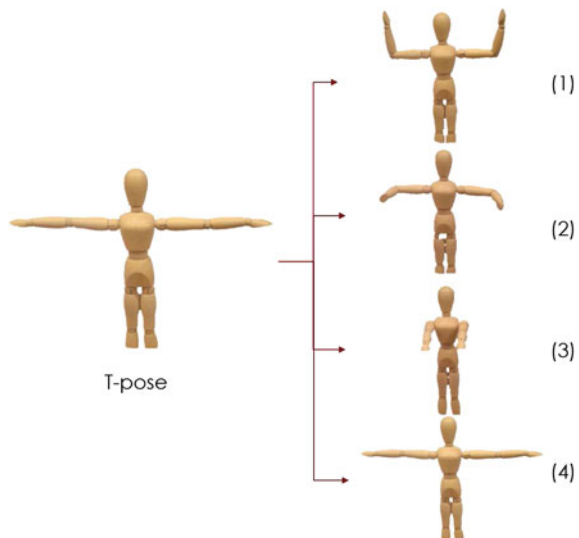
As a reference, measurements obtained from the optical multi-camera VICON motion capture has been used. Such systems are broadly used in the industry as well as to track the motion for i.e. biomechanical research [14]. According to data sheet published by the producer, declared precision of this system is down to 0.5 mm of translation and 0.5° of rotation with refresh rate up to 250 fps [13]. Such accuracy allows to use such measurements as ground-truth data. In performed experiments, VICON system worked with refresh rate 100 fps.

Performed experiments examined the right hand joints (elbow, wrist) positioning precision and the angle between the arm and the forearm estimation (the angle in the elbow joint) during 4 different movement sequences (Fig. 10):

1. Elbow flexion up to an angle of 90° .
2. Elbow flexion forward to an angle of 90° .
3. Straighten the hand in front of a body.
4. Stand still for more than a minute.

Selected gestures comprised motions that might be considered as challenging especially for Kinect. Each movement sequence started from the initial T-pose and was performed multiple times and averaged. The proposed method was also compared with the Kalkbrenner method precision, which was implemented according to the description included in the article [6]. The position estimation accuracy for elbow and wrist joints as well as the angle measurement accuracy has been presented in Figs. 11, 12 and 13. The gray color was used for Kalkbrenner position-based method results and the orange for the author's, orientation-based method achievements.

Fig. 10 Movement sequences performed during tests



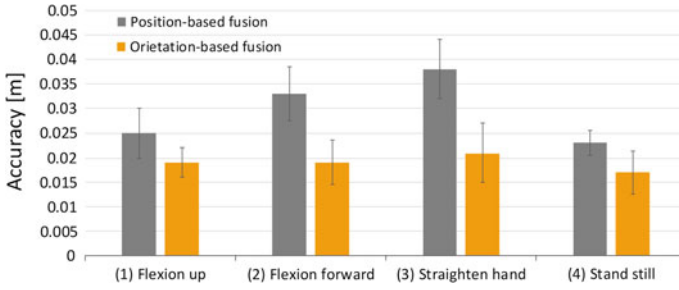


Fig. 11 Elbow positioning average accuracy

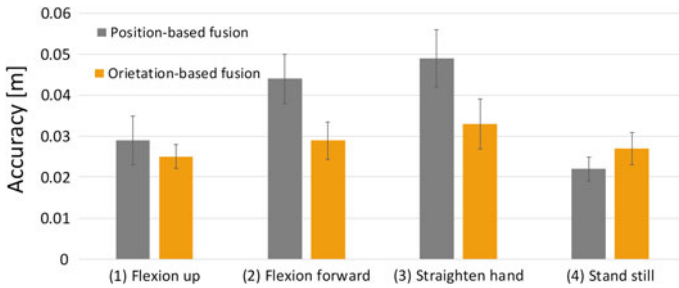


Fig. 12 Wrist positioning average accuracy

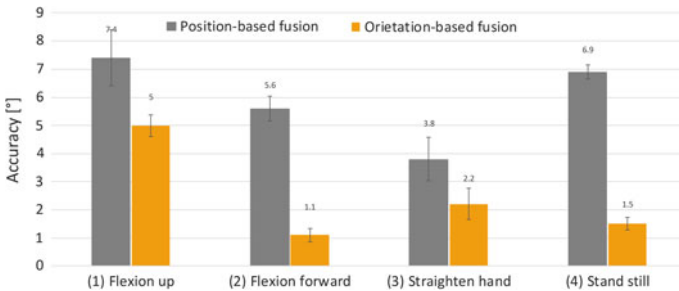


Fig. 13 Elbow angle measurement average accuracy

Unfortunately, during performed tests, authors were not able to achieve the described accuracy for the Kalkbrenner position-based method, however, the achieved results were close to the declared ones. Results presented on charts show the improvement in both: the position and the angle measurement accuracy. The average accuracy for the position estimation was 2.5 cm for the elbow and 2.9 cm for the wrist. The average angle measurement accuracy was 2.5°. The same values for position-based fusion estimations were 2.8 cm, 3.6 cm and 5.9° respectively.

5 Conclusion

The authors presented a new, orientation-based method for skeleton joints positioning. It was tested on variety of right hand movements. Obtained results have proven that the method managed to compensate imperfections of both measurement devices much better than previous approaches. Basing on the comparison of results gathered from the orientation-based fusion and the position-based fusion, the improvement of the estimation accuracy has been noticed and reached the rate of 15 up to 25 %.

Obtained results prove that the novel data fusion approach, based on the bones orientation, might be considered as an improved alternative to the well known, joint position-based methods.

References

1. Asteriadis, S. et al.: Estimating human motion from multiple Kinect sensors. In: 6th International Conference on CV/Computer Graphics Collaboration Techniques and Applications—MIRAGE '13. ACM, pp. 3–8 (2013)
2. Bo, A.P.L. et al.: Joint angle estimation in rehabilitation with inertial sensors and its integration with Kinect. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 3479–3483 (2011)
3. Destelle, F. et al.: Low-cost accurate skeleton tracking based on fusion of Kinect and wearable inertial sensors. In: EUSIPCO, pp. 371–375 (2014)
4. Feng, S., Murray-Smith, R.: Fusing kinect sensor and inertial sensors with multi-rate Kalman filter. In: IET Conference on Data Fusion and Target Tracking 2014: Algorithms and Applications, pp. 1–8 (2014)
5. Helten, T. et al.: Real-time body tracking with one depth camera and inertial sensors. In: 2013 IEEE International Conference on CV, pp. 1105–1112 (2013)
6. Kalkbrenner, Ch. et al.: Motion capturing with inertial measurement units and kinect—tracking of limb movement using optical and orientation information. In: Proceedings of the International Conference on Biomedical Electronics and Devices, pp. 120–126 (2014)
7. Khoshelham, K., Elberink, S.O.: Accuracy and resolution of kinect depth data for indoor mapping applications. *Sensors* **12**, 1437–1454 (2012)
8. Madgwick, S.O.H.: An efficient orientation filter for inertial and inertial/magnetic sensor arrays. Rep. x-io and University of Bristol (UK) (2010)
9. Obdrzalek, S. et al.: Accuracy and robustness of kinect pose estimation in the context of coaching of elderly population. In: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 1188–1193 (2012)
10. Pranay, P.: Human Male Sculpt (Maya and Zbrush Pipeline). <http://graphy-x-medley.blogspot.com/2013/07/human-male-sculpt.html>. Accessed 01 Jan 2016
11. Renaut, F., Nikolic, J.: MEMS Inertial Sensors Technology. Swiss Federal Inst. of Tech. Zurich, Autonomus Sys. Lab, Studies on Mechatronics (2013)
12. Tian, Y., et al.: Upper limb motion tracking with the integration of IMU and Kinect. *Neuro-computing* **159**, 207–218 (2015)
13. Vicon: Bonita Motion Capture Camera/VICON. <http://www.vicon.com/products/camera-systems/bonita>. Accessed 12 Feb 2016
14. Vicon: Case Studies/VICON. <http://www.vicon.com/case-studies>. Accessed 12 Feb 2016
15. Wan, E.A., van der Merwe, R.: The unscented Kalman filter for nonlinear estimation. In: Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium, pp. 153–158 (2000)

16. Wikipedia Contributors: High-pass filter—Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/wiki/High-pass_filter#Algorithmic_implementation. Accessed 01 Jan 2016
17. Woodman, O.J.: An Introduction to Inertial Navigation University of Cambridge, Computer Lab Technical Report, vol. 696, pp. 1–37 (2007)
18. XSense: MTx—Products—Xsens 3D motion tracking. <https://www.xsens.com/products/mtx/>. Accessed 10 Nov 2015