Hybrid method of human limb joints positioning hand movement case study

Grzegorz Glonek and Adam Wojciechowski

Lodz University of Technology, Lodz, Poland, grzegorz@glonek.net.pl, adam.wojciechowski@p.lodz.pl

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 sensors derived joints position fusion, mainly described in literature, with sensors derived bones orientations fusion and subsequent joints positions estimation. The paper comprises also the method evaluation results. It was veri ed both against professional motion tracking VICON system and Kalkbrener method [6], the most relevant to presented solution.

# 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, has even gained its separate name: 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– home. At the same time nonprofessional solutions reveal several imprecisions and 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 (30 Hz for Kinect and 70 Hz 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 [1,2,5] have proved that Kinect and IMU data fusion assures limbs joints positions tracking accuracy of about 5 cm, presented method achieves better results 3 cm.

# 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 [8]. Moreover, while tracking, lost joints may a ect the tracking accuracy of those which are fully visible to the sensor’s camera. The rotation of user body might be an example of such scenario. Basing on author’s experiments, if the user rotates more than 50 degrees (angle between 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 [9]. Figure 1 shows how estimated accuracy errors change with a distance. As the most important IMU aws, the gyroscope measurement drift and the temperature related bias in accelerometer measurements may be indicated [10,11]. The temperature inuence on bias is presented in a gure 2.

Presented papers estimate Microsoft Kinect general posture estimation variability is in range 5cm 10cm with variable bones length with accuracy 2cm-5cm

[12].

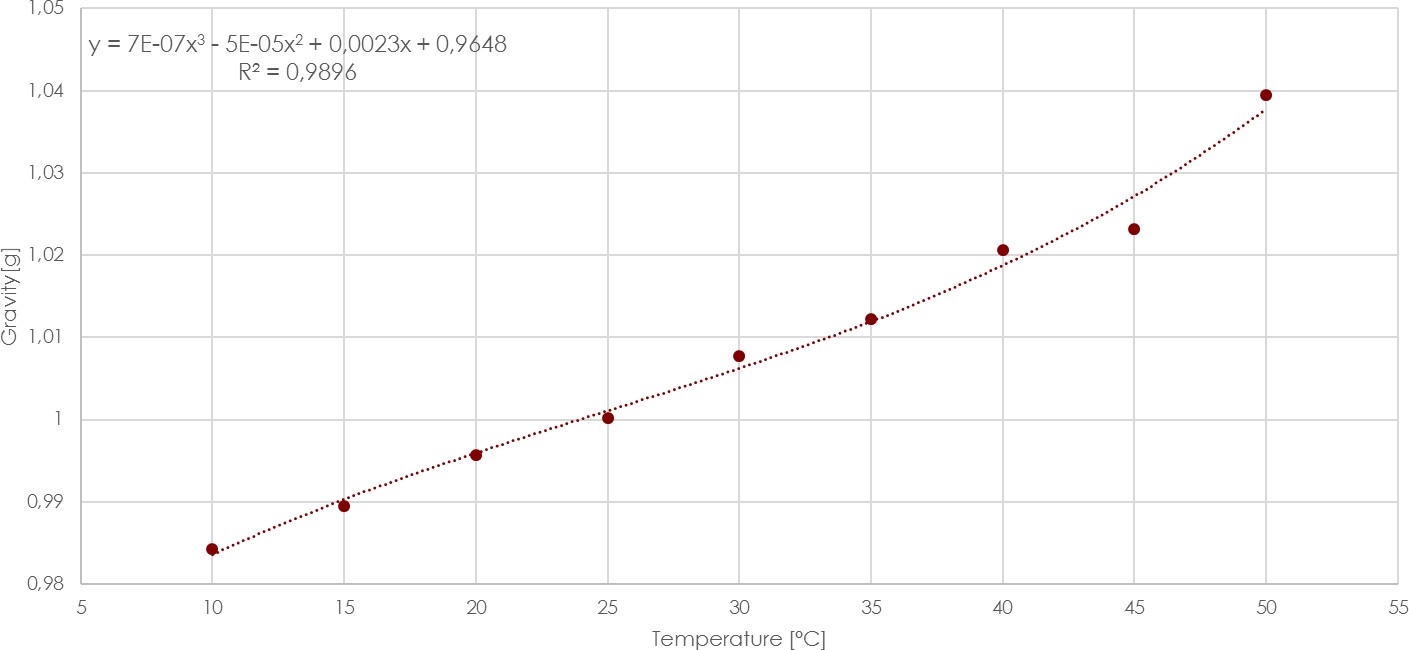
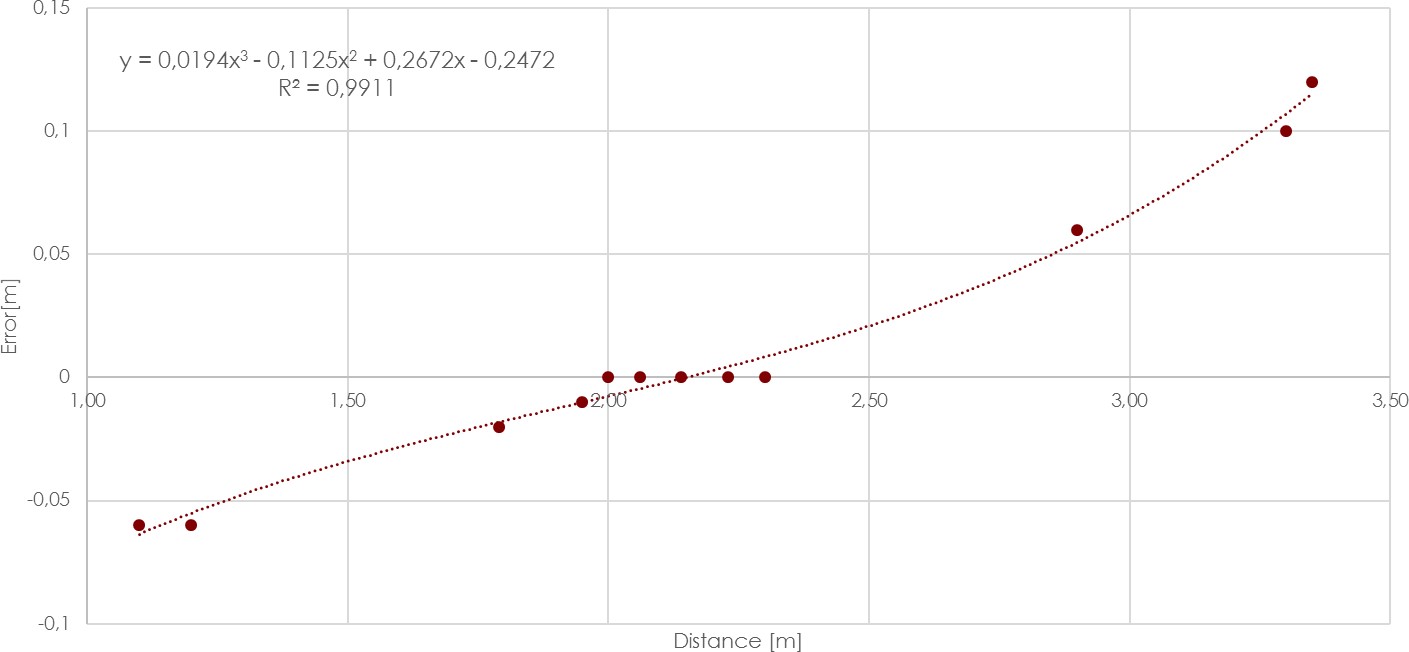


Fig.1. Position measurement error in Z-axis to distance from the Kinect

Fig.2. Gravity force measurement in temperature range 10◦*C*−50◦*C*

Considering selected controllers several hybrid data fusion approaches improving tracking accuracy can be found in literature. Authors have elaborated di erent approaches to fuse the sensors’ data which characterize various level of measurements reliability. First group of approaches can be classi ed as methods which use Kinect measurements as a reference system and relay on its measurements.

Bo et al. [1] described the joint angle (angle between joint adjacent bones) estimation method exploiting 5 degrees of freedom (DOF) IMU and Kinect. The method initial stage interprets gyroscope and the accelerometer data separately and their fusion by the linear Kalman lter (KF) was subsequently performed. The same angle estimated with Kinect data was used to initially calibrate and then temporarily correct the bias of accelerometer estimation. There were no numerical results provided in the paper however presented charts suggest data fusion resulting considerable improvement.

A di erent approach was presented in a paper published by Destelle et al. [2]. Authors decided to use set of 6DOF IMUs supported by magnetometer where each unit was sticked to one of the tracked limb bones. Basing on gathered measurements, orientation of each bone was estimated by Madgwick’s algorithm [4] and their superposition resulted in the full skeleton model. Kinect data was used twofold. The rst stage was to get the initial, reference skeleton frame to label data, estimated from inertial units and improve the IMU calibration. That process resulted in hierarchical de nition of bones orientations ( inertial skeleton ). The second stage of Kinect exploitation was to track position of central body point (torso joint) and then update the whole inertial skeleton relatively to this point displacement. Resulting, Vicon referenced knee joints angle estimation error varied between 4◦ an 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 quali ed to this group is the one presented in 2015 by Tian et al. [13]. Authors included in estimation process geometrical constraints of performed motion to eliminate estimations that are impossible to achieve in real life i.e. angle between forearm and arm greater that 180◦. Fusion algorithm used by authors based on Unscented Kalman Filter (UKF)[14]. Presented results shows that algorithm is able to work also while Kinect is outage which is not obvious in previously described methods. To estimate method accuracy, authors compared elbow angle measurement. Authors published information that angle measurement estimated by their fusion method deviates from the expected value for less than 20 degree.

Another group of published methods base on the assumption that both measurement devices imperfections need to be corrected continuously by the signal fusion. In 2014 Feng and Murray-Smith [5] proposed multi-rate Kalman lter based fusion method of joints positions estimated by Kinect with linear acceleration and velocity of this joint. Presented results shows that this approach stabilizes measurements around real value much faster that single rate KF. This is visible especially when movemnt starts/stops ( gure 3). The accuracy of presented method can be estimated around 1,5 2cm (based on published diagrams). However presented results refer to very short time periods (up to 5s) so it is impossible to estimate how method works in a long time.

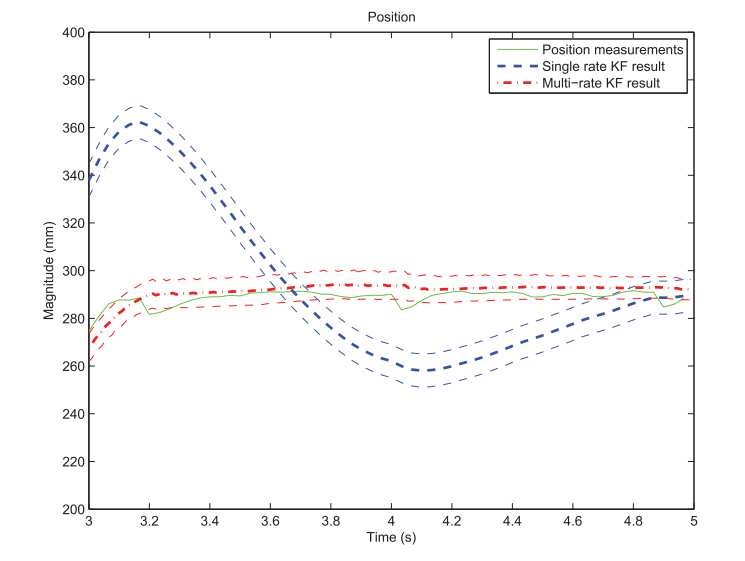


Fig.3. Position estimation with the multi-rate Kalman lter and the single rate Kalman lter[5]

A di erent 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 skeleton model (bones length) and Kinect measurements. Absolute joints positions estimation results were around ±2*.*5*cm* and seem to be the most accurate in long term experiments.

The approach presented in 2013 by Thomas Helten et al. [7] is more advanced than methods presented so far. 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 the 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 [17] which is a full set of professional inertial motion tracking system. The proposed method is based on the idea of visibility model from poses estimated on inertial and Kinect and then match them with prede ned poses stored in the database of 50 000 prede ned poses. However, such approach seems to be useless in scenarios where you need to estimate joints positions and other limb features. Our method is the most relevant to the Kalkbrenner approach but our main contribution is better, weighted sensor contextual in uence which results in overall better absolute joints positions estimations. The method joints absolute positioning precision of about 2-2.5 cm was veri ed against the ground-truth Vicon system.

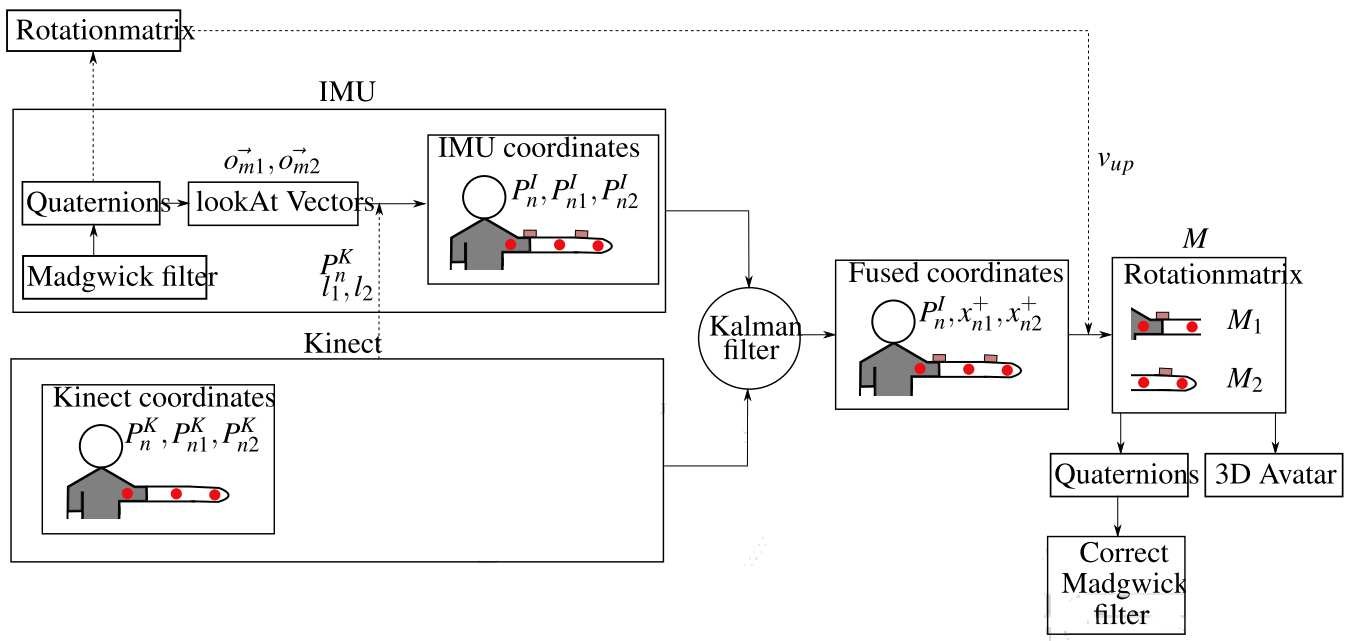
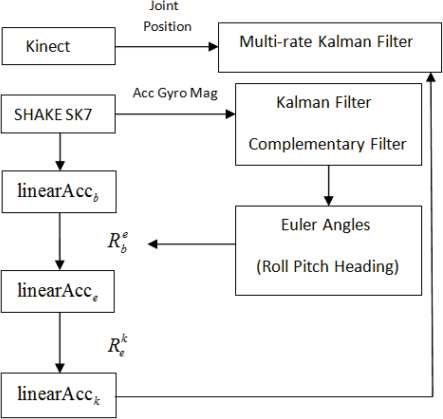


Fig.4. Feng and Murray-Smith [5] method diagram

Fig.5. Kalkbrenner et al. [6] method diagram

# 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 tracking method can be de ned as a function *f*:

 (1)

where:

*A* accelerometer measurement

*G* gyroscope measurement

*T* temperature measurement  start and end bone joints positions measured by Kinect i.e. elbow

and wrist

*QK* bone orientation estimated by Kinect *t* current time frame

*∆t* 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

) and bones orientations () supplied

by the Kinect device as well as accelerometer (*A* = [*ax,ay,az*]), gyroscope (*G* = [*gx,gy,gz*]) and temperature (*T*) measurements from each IMU. Kinect joints positions and IMU based marker locations on tracked limbs are presented in gure 3.

In the proposed method, data gathered from measurement devices, are denoised in the rst step and then used to calculate bones orientations. Then, orientations calculated from IMU devices and measured by Kinect are fused together and at the last step, bones length model is added to estimate absolute joints positioins. General overview of orientation-based fusion process is presented on gure 6.

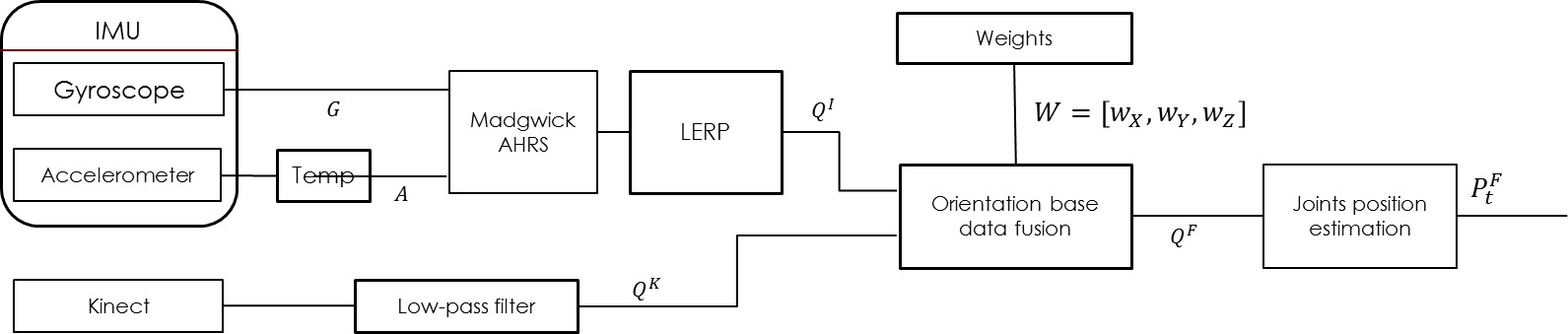
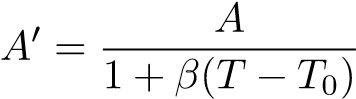


Fig.6. Orientation based fusion method

Data processing is performed in two parallel threads. The rst 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 rst thread, the IMU accelerometer bias was corrected with the equation 2, due to the device operating temperature destructive measurements in uence.

 (2)

where:

*A*0 corrected accelerometer measurement

*A* accelerometer measurement

*T* temperature measurement

*T*0 device reference operating temperature. For used device *T*0 = 25◦*C β* correction factor. For used device *β* = 0*,*0011

The value of *β* correction factor was the result of exploited IMU gravity regression analysis as a function of the device operating temperature ( gure 2). Next, the corrected accelerometer data and gyroscope measurements were used to calculate quaternion of adjacent bone orientation with Madgwick s lter [4]. The estimated orientation was then extrapolated to eliminate the observed delay. The linear interpolation algorithm was used in this step, represented with equation 3.

[*Φ,Θ,Ψ*]0 = [*Φ,Θ,Ψ*]*t* + *γ*([*Φ,Θ,Ψ*]*t* − [*Φ,Θ,Ψ*]*t*−1) (3) where:

[*Φ,Θ,Ψ*]0 corrected orientation in the form of Euler angles

[*Φ,Θ,Ψ*] orientation in the form of Euler angles

*γ* interpolation factor. For used device *γ* = 0*,*5

In the second thread, Kinect data needed to be denoised without the signi cant delay in measurements. It was done by rst-order exponential low-pass lter de ned by equation 4. Both joint positions and orientations has been ltered in this step.

*yt* = *αxt* + (1 − *α*)*yt*−1 (4)

where:

*y* ltered data *x* noised data *α* ltration factor. *α* = 0*,*065

The ˛– factor value has been estimated as a result of the analysis of the average Kinect positioning error during hand motion sequence (presented in gure 7).

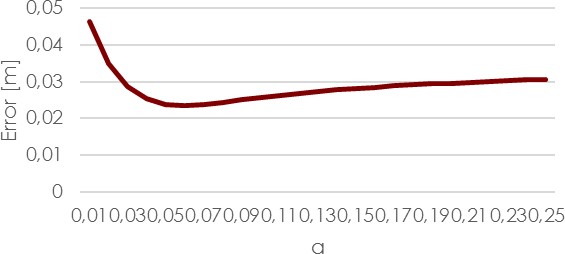
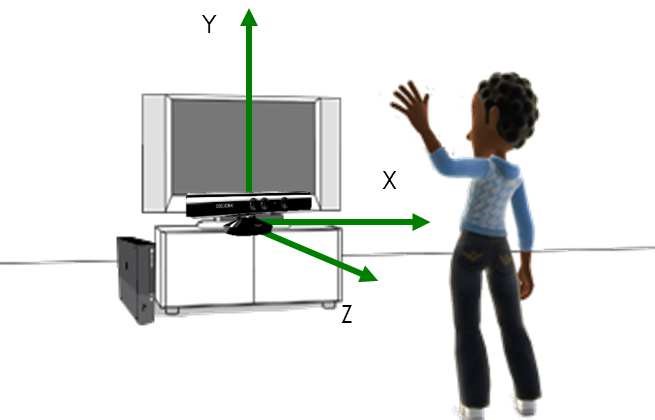
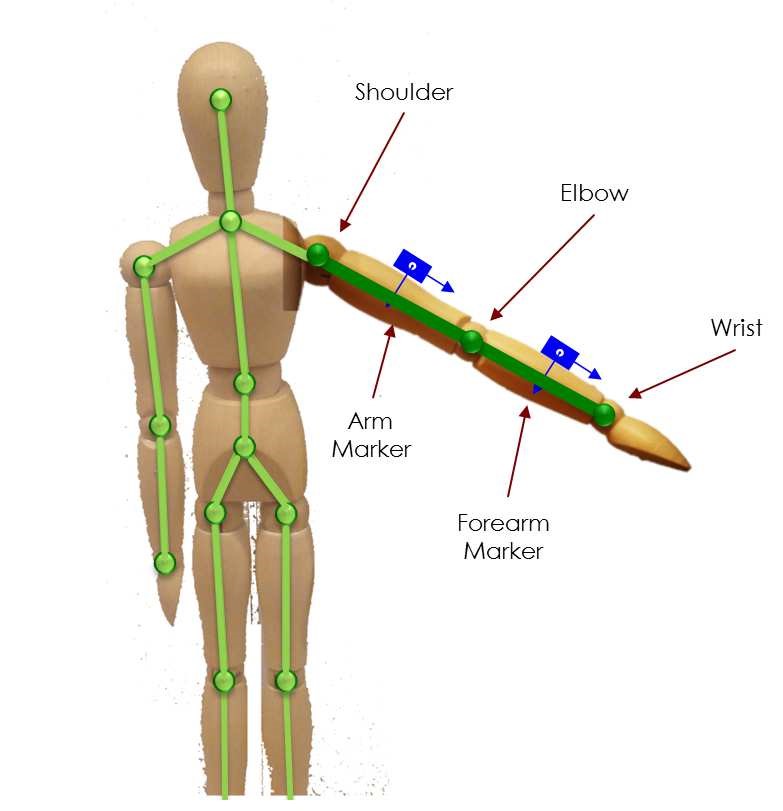


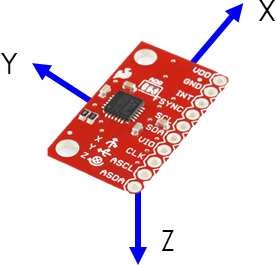
Fig.7. Kinect measurement accuracy to low-pass lter factor *α*

Both devices work in di erent coordination spaces ( gure 9) 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 has been chosen as the main one. That minimizes additional computations that need to be done.



(

a) Kinect



(b) IMU

Fig.8. Kinect skeleton joints positions

and IMU location Fig.9. Devices coordination spaces

In the next step, the controllers’ quaternions fusion was performed. It started with the judgment 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 lter in the form of equation 5 and sample results for keeping hands without motion and when Kinect lost tracking is presented on gure 10.

*nt* = *δnt*−1 + *δ*(*Pt* − *Pt*−1) (5)

[20] where:

*n* noise level *P* joint position *δ* ltration factor. *δ* = 0*,*01

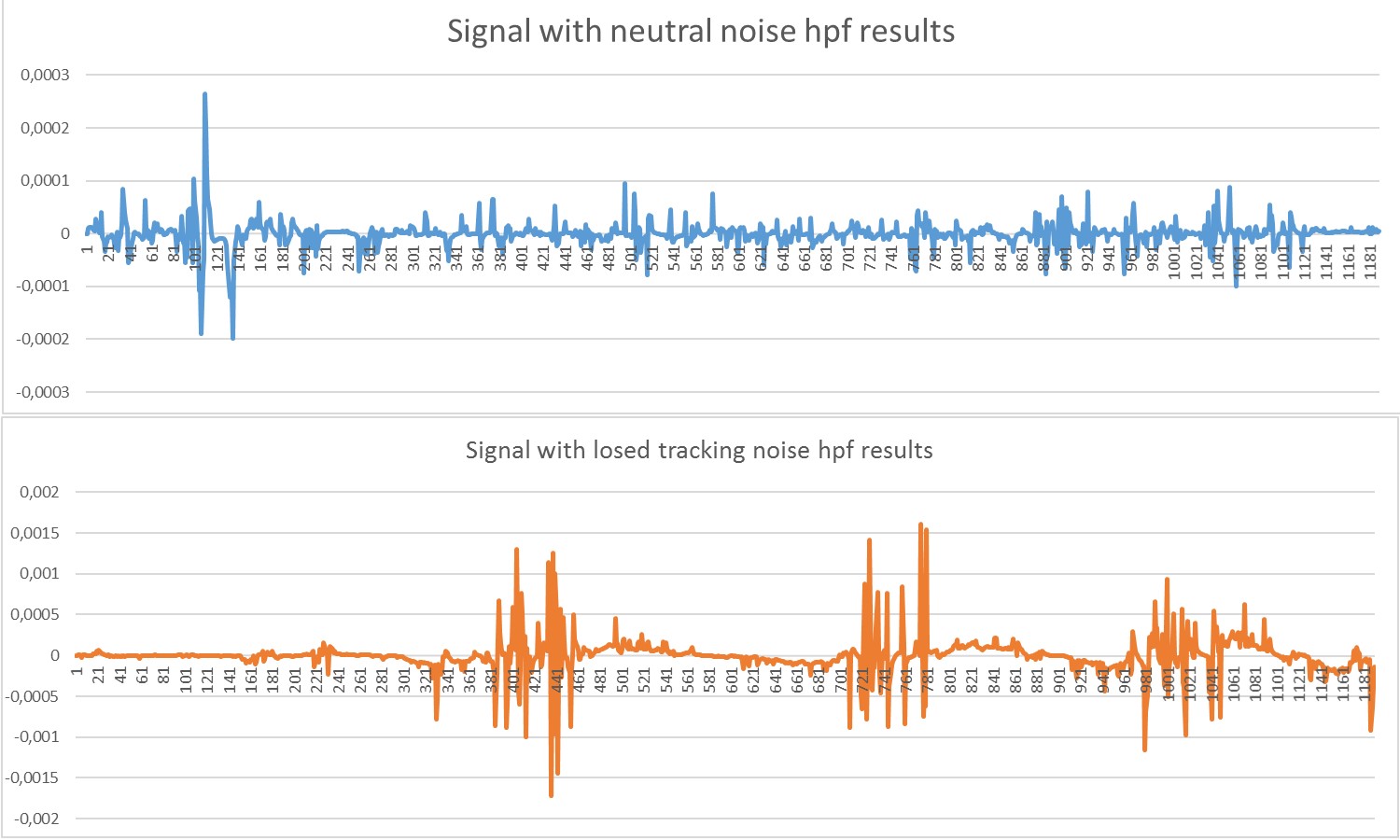
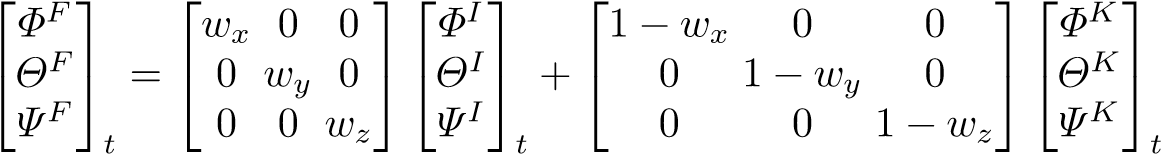


Fig.10. Kinect joint position measurement noise calculated with high-pass lter (eq.

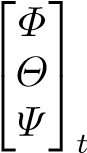
5)

If the user is rotated to Kinect camera more than 50◦ (the angle between line of user’s shoulders and the camera surface) or the noise level is too high (*>* 0*,*0004 based on performed experiments) then Kinect measurements are classi ed as unreliable and are replaced with IMU based orientation estimation change. The orientations fusion is de ned as the complementary process where rotations around each axis are joined with di erent weights. This approach was motivated by the fact that the controllers have di erent measurement abilities and accuracy in each axis.

If current Kinect s data is classi ed as reliable, the fusion is expressed by the following equation (eq. 6).

 (6)

where:

Euler form-based orientation: F - fused, I - IMU, K - Kinect

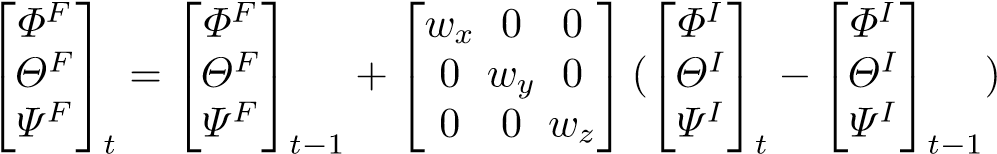
*wx,wy,wz* weights de ned fusion factor of each axis rotation. De nes as

[0*.*98*,*0*.*05*,*0*.*65] respectively.

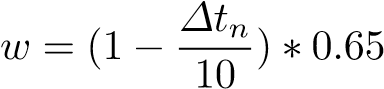
Weights used in equation 6 describes the level of importance of IMU measurements and need to be *<* 1. The higher value used, the more important measurement is. As Kinect is not able to measure rotation along ’x’ axis (Roll), weight close to 1 has been used. In case of usage inertial sensors without magnetic sensor support, rotation around gravity vector (’y’ axis Yaw) contains uncompensated time related drift so in this case IMU measurement is discriminated. Third axis rotation is measured by both devices however IMU have slightly better accuracy than Kinect which is re ected in weight *>* 0*.*5.

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

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

 (7)

In this case *wx* value remains the same and it equals 0*.*98 and *wy* and *wz* get low over the time according to the following function (eq. 8):

 (8)

*w* current weight value

*∆tn* 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, position of the desired joint might be calculated.

# Results

In order to verify 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 simultaneously with Kinect controller, two IMUs attached to hand arm and forearm bones and passive marker-based Vicon system. Markers were attached hand according to schema presented on gure 12 and were labeled as follow:

1. Wrist Top
2. Wrist Bottom
3. Wrist Back
4. Wrist Front
5. Elbow Top
6. Elbow Bottom
7. Elbow Back
8. Elbow Front
9. Shoulder Back
10. Shoulder Top
11. Shoulder Front

The PC used to record Kinect and IMU data was a 2.5 GHz Intel Core i74710HQ 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 ±500◦*/s* for gyroscope. Inertial 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.

Performed experiments examined the right hand joints (elbow, wrist) positions and the angle between the arm and the forearm (the angle in the elbow joint) during 4 di erent movement sequences ( g. 11):

1. Elbow exion up to an angle of 90◦
2. Elbow exion 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 of movement sequences started from the initial T-pose and were performed multiple times and averaged. The proposed method was also compared with the Kalkbrener method 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 gures 13, 14 and 15. The grey color was used for Kalkbrener, position-based, method results and the orange for the author’s, orientation-based, method achievements. Unfortunately, during performed tests, authors were not able to achieve the described accuracy for the Kalkbrner position-based method, however 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.

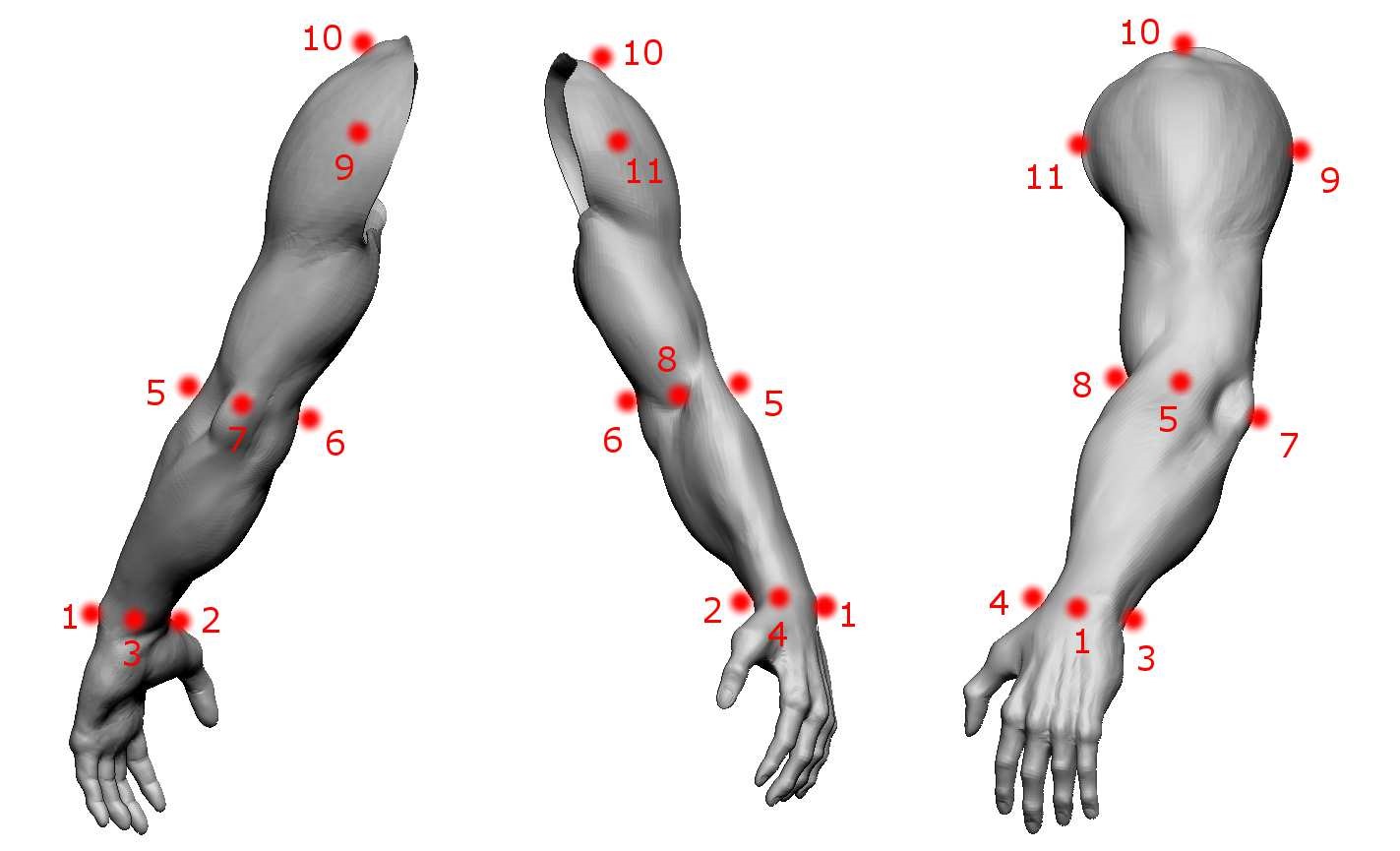
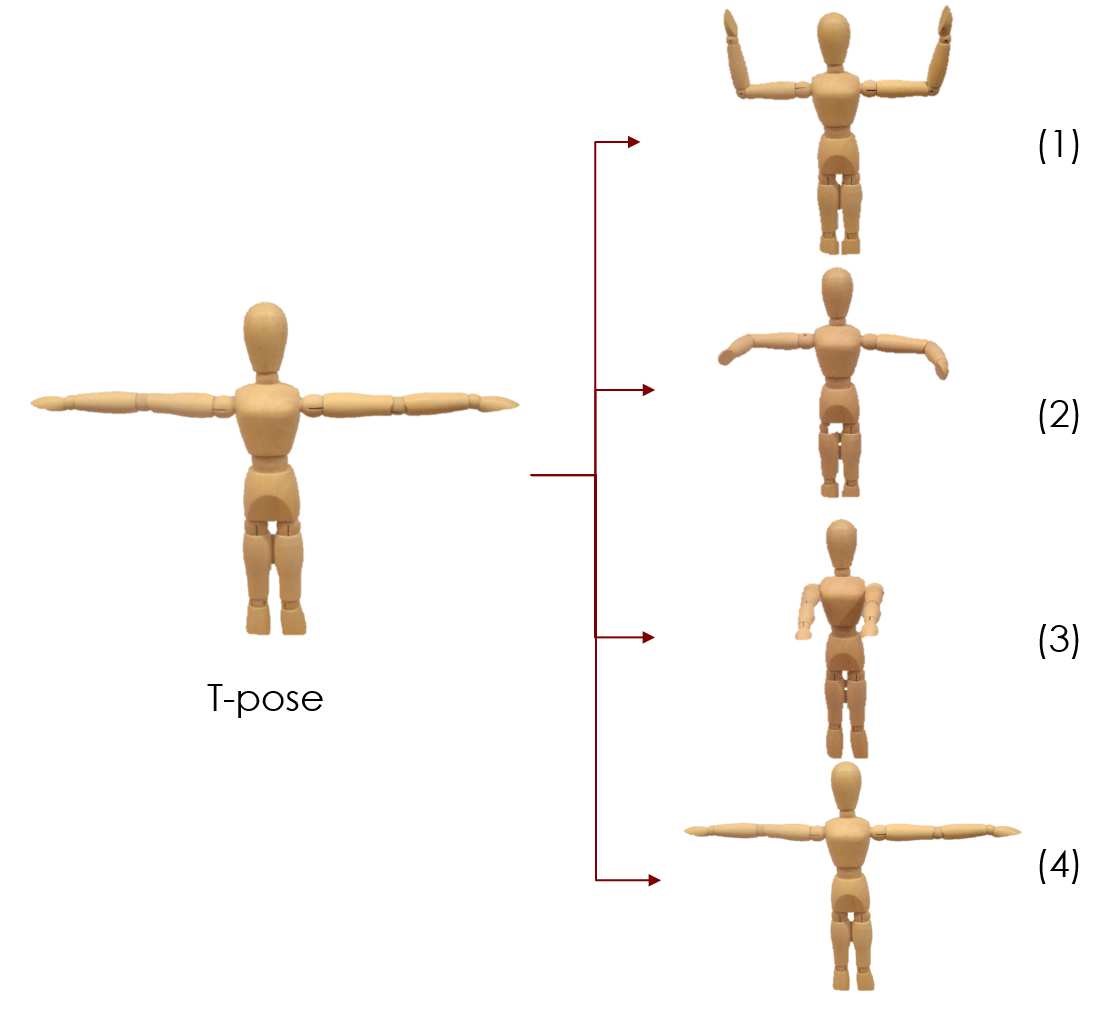


Fig.11. Movement sequences performed during tests.

Fig.12. Used Vicon Arm marker model.[[1]](#footnote-1)

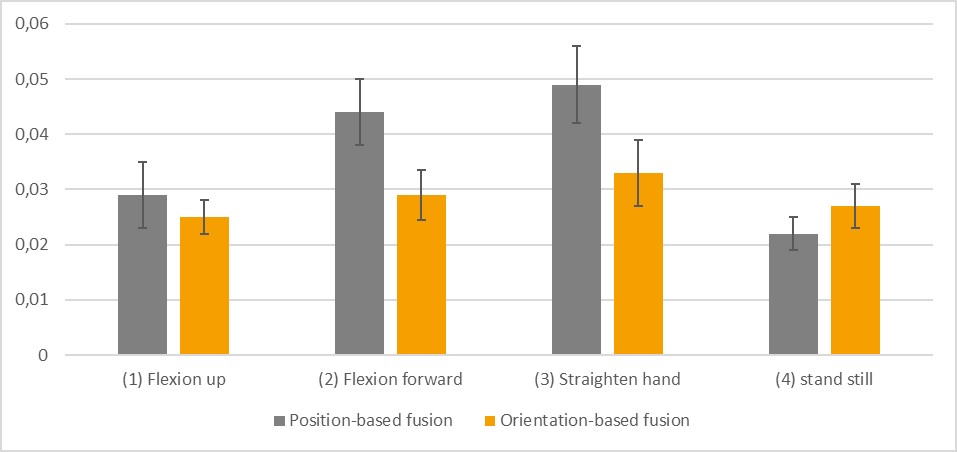
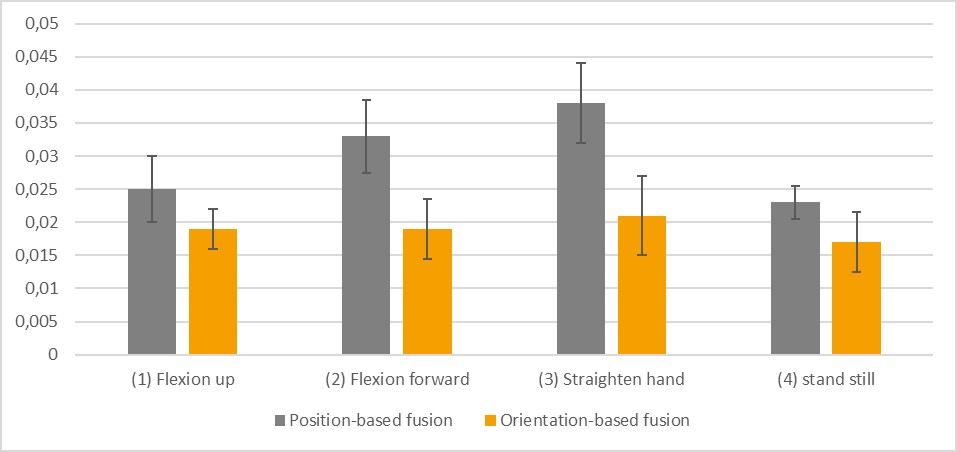


Fig.13. Elbow positioning average accuracy

Fig.14. Wrist positioning average accuracy

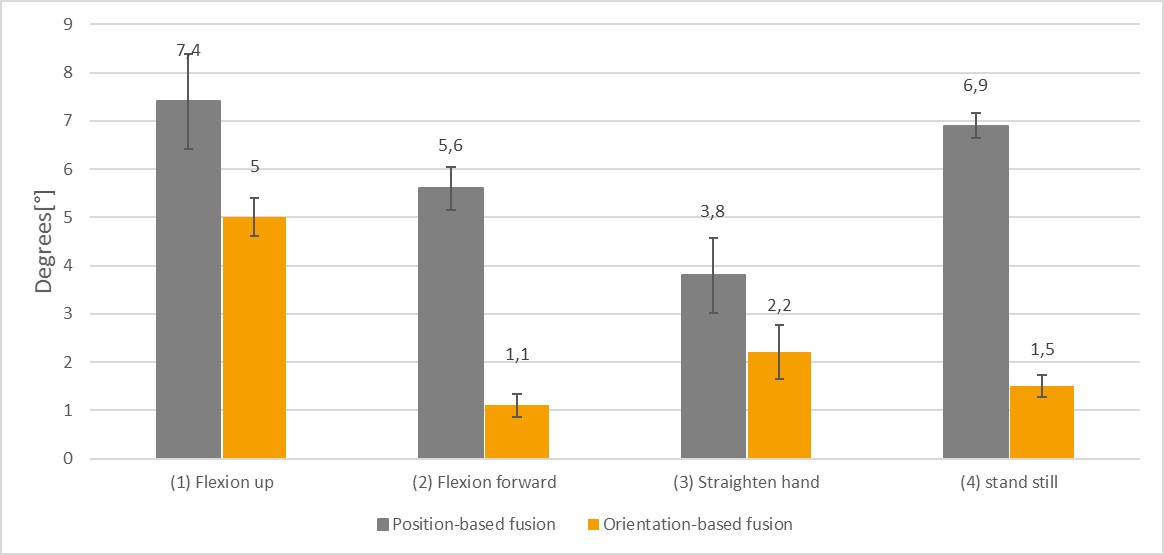


Fig.15. Elbow angle measurement average accuracy

# Conclusion

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

Obtained results proof 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. Bo, Antonio Padilha Lanari and Hayashibe, Mitsuhiro and Poignet, Philippe.: Jointangle estimation in rehabilitation with inertial sensors and its integration with Kinect. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS/2011 (2011)
2. Destelle, F. and Ahmadi, A. and O’Connor, N.E. and Moran, K. and Chatzito s,A. and Low-cost accurate skeleton tracking based on fusion of kinect and wearable inertial sensors.(2014)
3. Jean-Pierre Corriou: Process Control: Theory and Applications. Springer Science &Business Media (2013)
4. Madgwick, S.O.H.: An e cient orientation lter for inertial and inertial/magnetic sensor arrays. (2010)
5. Feng, Shimin and Murray-Smith, Roderick: Fusing Kinect Sensor and Inertial Sensors with Multi-rate Kalman Filter. IET Conf. Data Fusion Target Track. 2014 Algorithms Appl. (2014)
6. Kalkbrenner, Christoph and Hacker, Ste en and Algorri, Maria-elena andBlechschmidt-trapp, Ronald: Motion Capturing with Inertial Measurement Units and Kinect - Tracking of Limb Movement using Optical and Orientation Information. Proceedings of the International Conference on Biomedical Electronics and Devices (2014)
7. Helten, Thomas and Muller, Meinard and Seidel, Hans Peter and Theobalt, Christian: Real-time body tracking with one depth camera and inertial sensors. Proc. IEEE Int. Conf. Comput. Vis. (2013)
8. Asteriadis, Stylianos and Chatzito s, Anargyros and Zarpalas, Dimitrios and Alexiadis, Dimitrios S. and Daras, Petros: Estimating human motion from multiple Kinect sensors Proc. 6th Int. Conf. Comput. Vis. / Comput. Graph. Collab. Tech. Appl. MIRAGE ’13 (2013)
9. Khoshelham, Kourosh and Elberink, Sander Oude: Accuracy and Resolution ofKinect Depth Data for Indoor Mapping Applications. Sensors (2012)
10. Woodman, Oliver J.: An introduction to inertial navigation University of Cambridge Technical Report, Vol. 696 (20070
11. Renaut, Felix and Nikolic, Janosh: MEMS Inertial Sensors Technology (2013)
12. ObdrzÆlek, StepÆn and Kurillo, Gregorij and O i, Ferda and Bajcsy, Ruzena andSeto, Edmund and Jimison, Holly and Pavel, Michael: Accuracy and robustness of Kinect pose estimation in the context of coaching of elderly population. Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE (2012)
13. Tian, Yushuang and Meng, Xiaoli and Tao, Dapeng and Liu, Dongquan and Feng,Chen: Upper limb motion tracking with the integration of IMU and Kinect Neurocomputing (2015)
14. Wan, E.A. and Van Der Merwe, R.: The unscented Kalman lter for nonlinear estimation Proc. IEEE 2000 Adapt. Syst. Signal Process. Commun. Control Symp. (Cat. No.00EX373 (2000)
15. Walter, B.: Smartphonebasierte bewegungsanalyse zur therapiekontrolle bei morbus parkinson. Master s thesis, University of Applied Sciences Hochschule Ulm (2012)
16. Glaz, Y. L.: Smartphonebasierte bewegungsanalyse zur therapiekontrolle bei morbus parkinson. Master’s thesis, University of Applied Sciences Hochschule Ulm (2011)
17. XSense MTx Products Xsens 3D motion tracking.[Online]. Available: https://www.xsens.com/products/mtx/. [Accessed: 10-Nov-2015].
18. V. Ganapathi, C. Plagemann, D. Koller, and S. Thrun: Real time motion captureusing a single time-of- ight camera, IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 755 762 (2010)
19. Pranay P Pai Human Male Sculpt (Maya and Zbrush Pipeline) Available: http://graphyx-medley.blogspot.com/2013/07/human-male-sculpt.html [Accessed: 01-Jan-2016]
20. https://en.wikipedia.org/wiki/High-pass\_ lter#Algorithmic\_implementation[Accessed: 01-Jan-2016]

1. Arm’s render author: Pranay P. Pai. Source: [19] [↑](#footnote-ref-1)