

Quaterni-On: Calibration-free Matching of Wearable IMU Data to Joint Estimates of Ambient Cameras

Jochen Kempfle
Ubiquitous Computing
University of Siegen

Siegen, Germany
jochen.kempfle@uni-siegen.de

Kristof Van Laerhoven
Ubiquitous Computing
University of Siegen

Siegen, Germany
kvl@eti.uni-siegen.de

Abstract—Inertial measurement units, such as those embedded in many smartwatches, enable to track the orientation of the body joint to which the wearable is attached. Tracking the user's entire body would thus require many wearables, which tends to be difficult to achieve in daily life. This paper presents a calibration-free method to match the data stream from a body-worn Inertial Measurement Unit, to any body joint estimates that are recognized and located from a camera in the environment. This allows networked cameras to associate detected body joints from nearby humans with an orientation stream from a nearby wearable; This wearable's orientation stream can be transformed and complemented with its user's full body pose estimates. Results show that in multi-user environments where users try to perform actions synchronously, out of 42 joint candidates, our calibration-free method obtains a perfect match within 83 to 155 samples (2.8 to 5.2 seconds), depending on the scenario. This would facilitate a seamless combination of wearable- and vision-based tracking for a robust, full-body user tracking.

Index Terms—wearable inertial sensing, body pose capture, body joint tracking

I. INTRODUCTION

Wearable devices increasingly integrate inertial measurement units (or IMUs) and wireless transceivers, as both these technologies have matured into miniature, low-cost chips. This allows tracking of a user's physical movements for a myriad of tasks, such as gesture detection or activity recognition. This, however, is only possible for the body joint where this wearable is worn, for instance the lower arm in wrist-worn wearables. At the same time, breakthroughs in computer vision and depth cameras made it possible to detect human users and most of their individual body joints from cameras installed in the environment. These systems tend to work increasingly well but are limited to occlusion free environments and favorable lighting conditions. This paper presents a method that allows combining both modalities in real-time and without the need of camera-to-IMU calibration, by linking wireless data streams from IMU-based wearables to sets of joints recognized in a camera's field of view, as depicted in Figure 1.

Through the wireless capabilities of the wearables, it is possible to stream the orientation of the wearable, and thus the body joint that it is attached on, to cameras in the environment. These cameras tend to be able to capture multiple body joints in real-time, provided that those joints are not blocked by surrounding objects, people, or are not out-of-frame. These

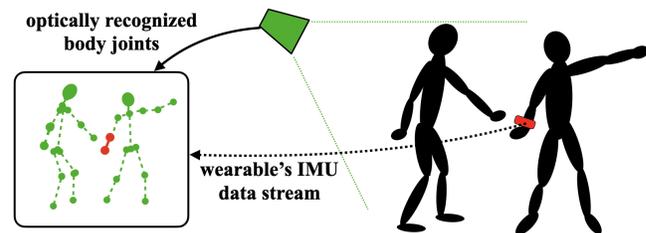


Fig. 1. Our method enables to match wireless streams of IMU data from a wearable device (red) to sets of body joints that have been optically tracked from an environmental camera (green). Once associated, the camera could then send back the user's full body poses to the right wearable as a service.

camera-based systems in the environment could return the full-body joint estimations back to the user's wearable as a service, thus making the wearable aware of its user's full body posture over time, without requiring users to wear a large number of devices. The body-worn devices on the other hand are not susceptible to occlusion or lighting conditions and can provide valuable data for a seamless body tracking or to resolve ambiguities. We envisage such complementing modalities to be extremely useful in applications that require full-body tracking, such as VR, character simulation, gesture controlled systems, or activity recognition. In the case of an unknown sensor setup, our method furthermore helps reducing the complexity for setting up such applications as it is able to automatically associate a body-worn device with its respective limb. A different, but promising application is indoor localization, where a person's position, as estimated from a camera system, can be forwarded to a wearable that was identified to be worn by that person.

II. RELATED WORK

Since the introduction of inertial sensors, several studies have been published that use inertial data to estimate where on the body these are likely attached. Early approaches to detect sensor placement from acceleration and gyroscope data of 4 or 5 sensors respectively during different activities [1], [2] have shown 100% accuracy for a walking activity, and up to 82% for several real-life activities. In [3], the authors further explored how sensor placement variations can influence human action recognition: It typically takes some time (up to a few minutes) to reach the peak accuracy, studies focusing on the sensor placements on head, wrist, torso, left breast pocket, and

front and back trouser pockets. A variety of classifier-based approaches have been proposed to see where inertial sensors are attached to: [4] uses a Support Vector Machine (SVM) to identify the location of 10 accelerometers on various parts of the body with an accuracy of up to 89%. Converging times, however, are not given and the method can not distinguish left from right body locations. A more frequent use of the right arm is assumed. In [5], a decision tree is trained on 17 inertial sensors placed on different limbs and achieves an accuracy of 97.5% in estimating the sensor placement. The setup requires a known sensor configuration and a walking pattern with sufficient arm movement (one participant with insufficient arm movement was excluded). Without knowing the sensor configuration, the accuracy drops to 75.9%. In [6], the sensor alignment and assignment on the lower body is estimated using deep learning. An accuracy of 98.57% is reported on the assignment classification using synthetic and real acceleration and gyroscope data for training. In [7], walking and non-walking accelerometer data from 33 participants, each wearing 5 accelerometers at ankle, thigh, hip, arm, and wrist, was recorded and the placement of each sensor estimated. Estimation was done through splitting the data in non-overlapping 10s windows and finding a walking motion with a SVM. If walking was detected, the location of the sensors is classified in a second step. Overall, a classification accuracy of up to 96.3% is reported using a majority voting strategy. Other works ([8], [9]) estimate the acceleration of feature points in a RGB or RGB-D stream and compare it to the acceleration readings of an accelerometer attached to a limb or an object to identify its location in the image domain.

As wearable-embedded inertial sensors evolved into Inertial Measurement Units (or IMUs), combining internal 3D accelerometer, gyroscope and magnetometer data into quaternions that reflect more accurate orientation, IMUs have been proposed to track the user's body pose (e.g. [10], [11], [12]). Orientations of most body joints can also be obtained from external cameras [13]. In this paper, instead of using accelerometer or gyroscope data as in previous work, we propose a novel method that uses the produced quaternions by two very different systems: The wearable's IMU data stream, and the potential body joints detected from an external camera. Matching these two, however, requires a set of steps as this process is far from trivial.

III. METHOD

In the following, we describe the preprocessing steps, as well as hurdles when matching a wearable IMU's quaternions to those coming from an external camera, after which we present four different comparison measures to identify the distance from the IMU orientation to the many joint orientations of multiple persons, as observed from a camera.

A. Preprocessing and Calibration Considerations

Both, a camera's and an IMU's orientations, are expressed in their distinct reference frame where the camera typically uses its view plane and pixel coordinates with the y-axis being the

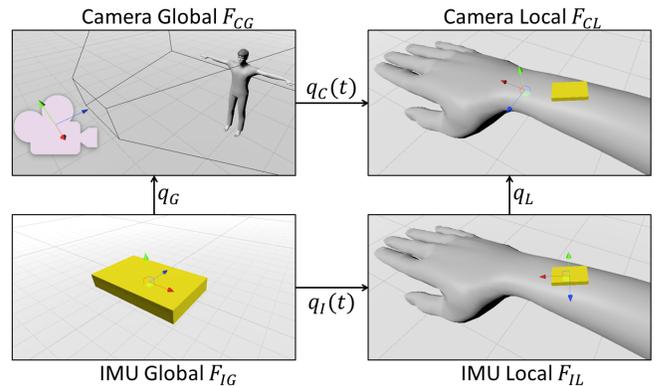


Fig. 2. From the top left to the bottom right: The camera's global and local, and the IMU's global and local coordinate systems. Transitions between these coordinate systems are marked with arrows and can be computed by multiplying with the respective quaternion q .

up-vector, and where the IMU typically uses the geomagnetic north and gravity vector as reference with the z-axis being the up-vector. A rotation around the up-direction therefore results in different orientations, depending on the device's reference frame. A simple and fast way to obtain an improvement in calibration, given the devices' coordinate mappings are known, is to shift the measured orientations to a common coordinate system by remapping their quaternions' bases such that they agree upon the axes of rotation. This step is not necessarily required for our proposed method, but enables the use of simple quaternion distance metrics and subsequently a comparison with our method. Since both reference frames still are not guaranteed to match, as will be elaborated in the following, the simple methods are limited to almost aligned reference frames. In a second step, the wearable IMU data is down-sampled to match the camera's sampling rate, using spherical linear interpolation.

As mentioned above, the camera and the IMU sensors do not share a common coordinate system. Fig. 2 illustrates the four different reference frames to be considered: The global and local camera frames F_{CG} and F_{CL} , and the global and local IMU frames F_{IG} and F_{IL} . Similarly there exist four distinct transitions between these. Transitions $F_{CG} \xrightarrow{q_C(t)} F_{CL}$ and $F_{IG} \xrightarrow{q_I(t)} F_{IL}$ define how to get from the camera or IMU global reference frame to the respective local camera or IMU frame at time t and are obtained as quaternion measurements $q_C(t)$ and $q_I(t)$. The other transitions are defined as $F_{IG} \xrightarrow{q_G} F_{CG}$ to get from the IMU global to the camera global reference frame, and as $F_{IL} \xrightarrow{q_L} F_{CL}$ to get from the IMU local to the camera local frame respectively. The quaternion q_G denotes the constant orientation offset of the global camera frame to the IMU frame and accounts for the fact that the camera can arbitrarily be placed in the environment, whereas q_L denotes the rotational alignment offset of the IMU when placed on a limb, given the camera can estimate the true limb orientation. To switch between the different reference frames, consider the equality of the transitions $F_{IG} \xrightarrow{q_I(t)} F_{IL}$ and $F_{IG} \xrightarrow{q_G} F_{CG} \xrightarrow{q_C(t)} F_{CL} \xrightarrow{q_L} F_{IL}$. Expressed as

quaternion equation, we get (1) with \circ denoting the quaternion or Hamilton product.

$$q_I(t) = \overline{q_L} \circ q_C(t) \circ q_G \quad (1)$$

Both, q_G and q_L , usually are not known and need to be determined through calibration, also known as the hand-eye calibration problem in robotics. To solve the calibration problem, we need a series of measurements $q_I(t)$ and $q_C(t)$. The transition from $q_I(t)$ to $q_I(t+i)$ ($i \in \mathbb{N}^+$) is given as $F_{IL} \xrightarrow{q_I^{-1}(t)} F_{IG} \xrightarrow{q_I(t+i)} F_{IL}$. Similarly, using the camera measurements $q_C(t)$, we have the path $F_{IL} \xrightarrow{q_L} F_{CL} \xrightarrow{q_C^{-1}(t)} F_{CG} \xrightarrow{q_G^{-1}} F_{IG} \xrightarrow{q_G} F_{CG} \xrightarrow{q_C(t+i)} F_{CL} \xrightarrow{q_L^{-1}} F_{IL}$. The camera offset q_G cancels out and we obtain (2).

$$q_I(t+i) \circ \overline{q_I(t)} = \overline{q_L} \circ q_C(t+i) \circ \overline{q_C(t)} \circ q_L \quad (2)$$

Substituting $q_A = q_I(t+i) \circ \overline{q_I(t)}$, $q_B = q_C(t+i) \circ \overline{q_C(t)}$, and $q_X = \overline{q_L}$ yields the (reordered) calibration equation (3).

$$q_A \circ q_X = q_X \circ q_B \quad (3)$$

Finding the IMU limb offset $q_X = \overline{q_L}$ using a series of n different observations now is subject to:

$$\arg \min_{q_X} \sum_{n \in \mathbb{N}^+} \|q_{A,n} \circ q_X - q_X \circ q_{B,n}\| \quad (4)$$

Explicitly solving (4) does not yield a match of both modalities, but for each IMU a specific IMU to limb offset estimate q_X with an unclear outcome for non-matching orientation streams. It is susceptible to noise and unstable orientation estimates and furthermore, it is a very expensive operation that would have to be repeated for every IMU for every observed joint in camera space. Instead, we aim for a calibration free matching method that does not need to solve (4). With (3) we already are independent from the camera offset and only the IMU-limb alignment has to be considered as will be detailed in section III-B. After matching and if required, calibration only needs to be performed once per IMU. An algorithm to solve (4) can for instance be found in [14] from which we also adapted above calibration scheme.

B. Quaternion comparison measures

In this section, we introduce four different comparison measures that can subsequently be used to match quaternions. Assuming small camera-IMU and IMU-limb offsets, from (1) follows that $q_I \approx q_C$. A first, straightforward option thus is to use the **(1) Quaternion angle** or geodesic angle between two quaternions. It is computed with the dot product as in (5).

$$d_q(q_I, q_C) = 2 \arccos |\langle q_I, q_C \rangle| \quad (5)$$

As estimates of the limb joints from camera-based systems tend to be unstable and suffer from randomly swapping around the limb's direction, a second measure would be the **(2) Stable quaternion angle**. It makes use of the quaternion swing-twist-decomposition and only keeps the swing part around the limb

direction, in our case the x-axis. The swing quaternion then is forwarded to (5) to obtain (6):

$$d_{q,stable}(q_I, q_C) = d_q(\text{swing}(q_I), \text{swing}(q_C)) \quad (6)$$

Using the stable quaternion angle also is useful if the IMU's rotation around the attached limb is not known or when from the camera-based pose estimation only joint positions are available. The sensor's axis in the direction of the limb however has to be known.

Since the assumptions for small reference system offsets made in above metrics do not hold in general, we propose the **(3) Independent quaternion angle**. Reorganizing equation (3) to $q_A = q_X \circ q_B \circ \overline{q_X}$ and considering that the real parts at both sides need to be the same, yields:

$$\begin{aligned} \Re(q_A) &= \Re(q_X \circ q_B \circ \overline{q_X}) \\ &= w_x^2 w_B - w_x \vec{v}_X \cdot \vec{v}_B + w_B \vec{v}_X \cdot \vec{v}_B \\ &\quad + w_B \vec{v}_X \cdot \vec{v}_X + \vec{v}_X \times \vec{v}_B \cdot \vec{v}_X \\ &= w_B (w_X^2 + \vec{v}_X \cdot \vec{v}_X) = w_B = \Re(q_B) \end{aligned} \quad (7)$$

From (7) follows that the real parts of q_A and q_B are equal, which means that we have a way to discard q_X and thus only depend on the measurements. Intuitively, (7) can be understood as: the amount of rotation or the angle in between two successive measurements at times t and $t+i$ measured by both sensors needs to be the same, independent of the direction or axis of rotation of each. This makes sense since the limb rotation does not depend on the sensor alignment.

With $q_A = q_I(t+i) \circ \overline{q_I(t)}$ and $q_B = q_C(t+i) \circ \overline{q_C(t)}$, both representing the rotation from the respective quaternion $q(t)$ to $q(t+i)$, and considering that both comprise the same amount of rotation in between time points t and $t+i$, we can compute the angle between both using (5). The independent distance metric then is defined as:

$$d_{ind}(q_I, q_C) = |d_q(q_I(t), q_I(t+i)) - d_q(q_C(t), q_C(t+i))| \quad (8)$$

Considering stability issues of the camera-based limb orientation estimation, similar to the stable quaternion angle, we introduce the **(4) Independent stable quaternion angle**. It is based on (8), but instead only uses the swing component of a quaternion, as stated in (9).

$$d_{ind,stable}(q_I, q_C) = d_{ind}(\text{swing}(q_I), \text{swing}(q_C)) \quad (9)$$

C. Discrete Joint Matching

To find the body joint that was picked up by the camera and that matches the wearable's IMU orientation sequence best, first a distance matrix $d[k][n]$ for each camera joint k and sample n is computed using any of the four distance metrics described above. Given a distance matrix d , (10) then computes the most likely camera joint k the IMU is attached to at time t and within a window comprising w samples.

$$\text{match}(d, t, w) = \arg \min_k \frac{1}{w} \sum_{n=t}^{t+w} |d[k][n]| \quad (10)$$

Equation (10) allows to identify the limb position of an IMU at any time point independently, even if its body location has changed in the meantime.

IV. EXPERIMENT DESIGN

To validate our method, we collected a dataset in which three participants were simultaneously captured by a Kinect v2 depth camera while performing near-synchronized movements. One of the study participants was wearing an IMU device that delivered quaternions wirelessly to a system attached to the Kinect, in which also all the participants' joints are calculated from the depth data in real-time. The IMU was worn in two different constellations: (1) on the wrist, as one would wear a smart watch, and (2) in the user's pocket, as one might carry a smartphone. To make the task of estimating on which joint (of overall 42 optically detected joints) the IMU is worn particularly challenging, these three different scenarios with high synchronicity were chosen to evaluate the performance of the methods:

- (A) The *Macarena* line dance, in which participants tend to move one limb at a time, in a synchronous fashion. Participants were at the start of the recording only slightly familiar with the Macarena movements and started asynchronous, though they improved after a few repetitions through listening to and watching the music video as they performed the dance.
- (B) The *head, shoulders, knees and toes* exercise for children, causing participants to move their left and right limbs synchronously. Motion sequences are shorter for this scenario, and participants quickly became familiar with the few movement sequences for this exercise.
- (C) The participants *walking* along the room parallel to the camera's line of sight in a synchronous fashion. In this scenario, the participant with the wearable set the pace whereas the two others were trying to walk in the same pace and rhythm.

For scenarios (A) and (B), the IMU was worn on the right wrist with negligible IMU-limb offset, and for the walking scenario (C), the IMU was worn in the front left pocket with the IMU not being properly aligned to the limb. The camera-IMU offset was about 25° and the IMU was worn by the same participant in all scenarios. Recording times are (A) 95 s, (B) 45 s, and (C) 64 s.

Sensor systems. The wearable IMU that we used in the experiment is a custom wireless sensor module that is built around the Bosch BNO055 IMU, delivering the sensor's orientation as a quaternion at a sampling speed of 100 Hz. It can be used as a single sensor, or combined in a network of multiple IMUs, using Nordic Semiconductor's nRF24L01 low-power transceivers. It runs approximately for 18 hours continuously with a miniature 400 mA battery. For the estimation of the users' joints from the environment, a Kinect v2 framework is used as a well-known depth camera system that performs optical tracking of users' body joints through a method presented by Shotton et al in [15]. For the datasets generated in this paper's experiments, we have stored the detected body joints of all users, as quaternions, at a sampling rate of 30 Hz, thus sub-sampling the wearable's stream, to be able to focus in the results on the matching itself.

V. EVALUATION RESULTS

For the evaluation results below, equation (10) is evaluated for all distance measures (also see III-B). For the independent metrics (8) and (9), the time offset parameter is set to $i = 7$. At lower values, especially at $i = 1$, the estimation accuracy degrades as there is insufficient movement in between successive samples. An example visualization of the distance matrix from the stable quaternion metric is shown in Fig. 3. It nicely visualizes the dynamics of the joint distances caused by the rhythm of the Macarena line dance. The IMU hereby was attached to the "Body3/WristRight" joint and the corresponding joint classification is highlighted in magenta. Although locally other joints have a smaller distance, within broader windows it will overall have the closest distance to the IMU.

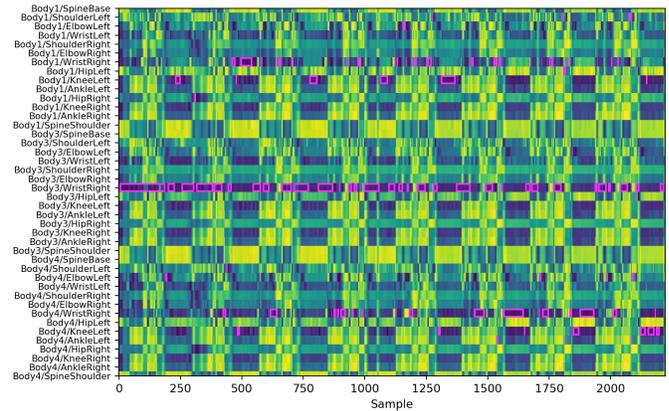


Fig. 3. Joint distances from stable quaternion distance, equation (6), of the Macarena line dance, with evaluated joints in the rows and samples in the columns. Blue being low and yellow being high distances. Joint classification using a window size of $w = 20$ samples is indicated with magenta edges.

Joint matching accuracy. To assess the performance of the different metrics, the window length w step wise is increased and for each w all samples are classified by moving the window over the respective distance sequence. The joint matching accuracy is computed as the amount of correctly classified window positions divided by the total number of window positions available for a certain w . Figure 4 shows the matching accuracy of all three scenarios against an increasing window length w of the moving window.

In scenario (A), the stable independent angle distance is the only metric that achieves 100% accuracy. It converges beyond window lengths of $w = 128$ samples (4.27 seconds) and already is close to 100% at $w = 96$ (3.2s). The stable quaternion metric reaches its maximum of 87% at $w = 195$ (or 6.5s) and the quaternion angle metric stays with small deviations at about 45% over all window lengths. The independent quaternion angle peaks with 17% at $w = 14$ and afterwards decreases to 0%. Since the Kinect v2 has difficulties in correctly estimating the twist orientation of wrist joints, with large angular offsets on successive samples, the methods that are not stable against such errors can not correctly assign the wrist-worn IMU to the matching joint. Especially the independent quaternion angle requires successive samples of both streams to have similar changes in rotation.

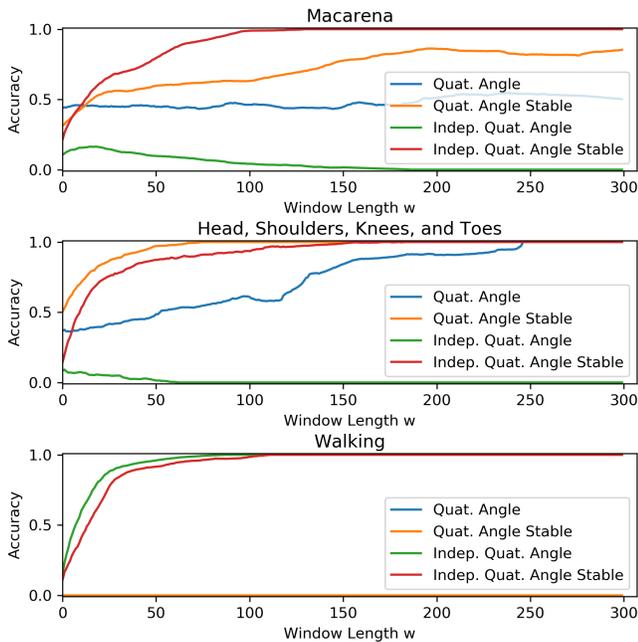


Fig. 4. Accuracy of the different matching methods, plotted for the different experiment scenarios with increasing window sizes w . The IMU sensor was worn on the wrist in the Macarena and Head, shoulders, knees, and toes scenarios. For walking, it was worn on the hip. The proposed *independent stable quaternion angle* metric can accurately match the correct joint in all scenarios. Depending on the scenario, other metrics nevertheless can show a slightly better performance.

A similar behaviour can be observed in scenario (B). The large errors in the camera's wrist orientation estimates cause the independent quaternion angle metric to match other joints that comprise smaller deviations around any axis, resulting in close to 0% or 0% accuracy at all window lengths. The stable independent quaternion angle is able to remove the twist rotation of the wrist and converges to 100% accuracy above a window length of $w = 155$ samples (or 5.2s). The stable quaternion distance metric in this scenario converges fastest to 100% accuracy above window lengths of about $w = 75$ samples (2.5s). The quaternion distance requires at least $w = 246$ samples (or 8.2s) to accurately match the correct joint. In contrast to scenario (A), here both the stable and normal quaternion angle metrics can more efficiently match the correct joint. One reason for this might be that during the Macarena line dance only one limb at a time is moved while during the head, shoulders, knees, and toes exercise many limbs are moved simultaneously and thus any ambiguities can be resolved within smaller time windows.

For the walking scenario (C), only the independent quaternion angle metric and its stable variant are able to correctly assign the IMU to the upper left leg. The first metric hereby converges faster to 100% accuracy at window lengths above $w = 83$ samples (2.8s), closely followed by the stable version at above $w = 108$ samples (3.6s). All other distance metrics can not at all match the IMU to the correct camera joint. The reason is that in this scenario the IMU is not well aligned to the limb and due to this offset any other random camera joint appears to be closer to the sensor orientation at any time.

IMU-to-camera offset. For the case in which the IMU-to-camera offset is not known in advance, we model the effect on the assignment accuracy by step-wise increasing the offset from -180° to 180° around the camera's up-axis, with 0° being the unchanged orientation. Figure 5 plots the accuracy of all activities against varying IMU-to-camera offsets at a fixed window length of $w = 50$ samples.

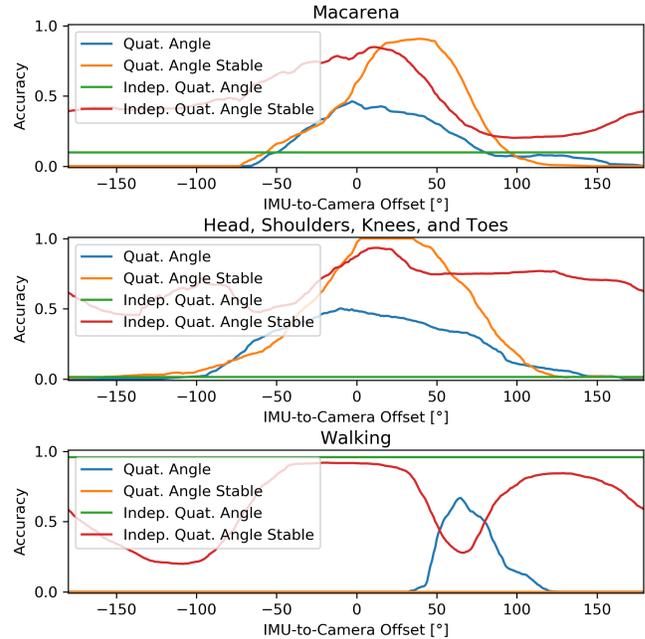


Fig. 5. Accuracy of the assignment with different IMU-to-camera offsets along the up-axis at a window length of 50 samples, for the three scenarios. Top: Macarena (A); Middle: Head, shoulders, knees, and toes (B); Bottom: Walking (C). For scenarios (A) and (B), the IMU was worn on the wrist, for scenario (C), the IMU was worn on the hip.

In scenario (A), the accuracy of the quaternion angle peaks with 47% at -3° . Its stable variant reaches its maximum of 90% in between 18° and 49° . Both metrics decrease to 0% to both sides. The stable independent quaternion angle shows at 10° an accuracy of 85% and has a decreased performance to both sides, however without dropping to 0%. As this metric relies on the swing-twist decomposition of the measured quaternions, it is not fully independent from the IMU-to-camera offset. The independent quaternion distance is not affected by any rotation offset, but has a low accuracy of about 10% due to the unstable wrist orientation estimates of the Kinect v2.

A similar behaviour can be observed for scenario (B). Here the quaternion angle metric peaks at -10° with an accuracy of about 50%, and the stable quaternion distance reaches its maximum of 100% between 3° to 35° . Both metrics drop to 0% to both sides. The independent quaternion angle again can not deal with the wrist error while its stable variant has its maximum accuracy of about 93% in the range of 7° to 22° and maintains an accuracy of above 45% at all other offsets.

In scenario (C), the quaternion angle metric peaks at 65° with an accuracy of 67% and drops to 0% at both sides. Its stable version remains at 0% for all camera offsets. We assume that the quaternion distance metric is at the 65° offset

more likely to match the correct joint due to comprising additional rotation information around the x-axis. The independent quaternion angle does not suffer from erroneous camera joint estimates and stays at an accuracy of 96% over the whole range of camera offsets, showing the advantage of truly being independent. Its stable version has its maximum accuracy of 92% between -43° and 25° . Two local minima of about 20% and 27% are between -127° to -98° and at 66° .

VI. DISCUSSION

Our dataset comprises many relevant challenges of IMU to camera joint matching, namely synchronous movements, erroneous joint orientation estimates, IMU-limb and IMU-camera misalignments, and asynchronous sampling rates. It, however, does not contain a scenario with regular occlusion events and it is with only three different scenarios, three participants and only one IMU per scenario somewhat limited to draw final conclusions about its robustness in real world environments. Especially in case of occlusions, its performance is likely to decrease significantly, but once a joint was successfully associated, a continuous tracking through both complementary modalities is facilitated. Both independent quaternion angle methods do not require calibration, but, in contrast to both other metrics, require user movement for the matching process in all circumstances. The independent quaternion distance metric, however, is susceptible against estimation errors. Its stabilized version can compensate for that, but is, due to the swing-twist decomposition being affected by calibration parameters, less robust against calibration offsets. All proposed metrics only require a few processor instructions and can be computed in parallel for all IMU-camera joint combinations, thus being highly performant even for large numbers of joints. The most important tasks for the future work are to evaluate the methods on a broader dataset, including more sensors as well as occlusion events, and to tackle the dependency of the swing-twist decomposition on calibration parameters.

VII. CONCLUSIONS

We presented a method that allows the quaternion stream from a wearable IMU sensor to be matched, on the fly, with quaternion estimates extracted from an optical sensor (e.g., a depth camera), thus allowing to track the user's full body posture over time. Our method accounts for different coordinate systems, as well as inaccuracies that tend to be present in optical body pose estimation frameworks (such as sudden twists in the estimates from the wrists).

We performed a series of experiments with participants performing synchronous dance routines, using 30 Hz depth cameras and body-worn IMU sensors. Results show that our method can find the matching joint of the correct user within 75 to 128 samples (or 2.5 to 4.3 seconds) at the wrist, using the stable or the independent stable quaternion metrics respectively. The independent stable metric overall is the better measure in scenarios (A) and (B) since it delivers optimal results in both. For walking, when the IMU is placed at the pocket of the upper leg, best results are obtained from any of

both independent quaternion metrics that find the matching joint within 83 to 108 samples (or 2.8 to 3.6 seconds). While the standard quaternion distance metrics may have their benefits in calibrated scenes, the calibration independent metrics have their big advantage in environments with unknown setups. Both the stable and normal distance measures have shown to have their specific area of application, depending on the stability of the joint orientation estimates.

The dataset and the source code for our method are publicly available at <https://ubicomp.eti.uni-siegen.de> for further development and to support reproduction of our results.

VIII. ACKNOWLEDGEMENTS

This research was partially funded by the DFG project LA 2758/8-1, PervaSafe, and by the BMBF GAIIA Project 01DG20001, AfricaSign.

REFERENCES

- [1] K. Kunze, P. Lukowicz, H. Junker, and G. Tröster, "Where am i: Recognizing on-body positions of wearable sensors," pp. 264–275, 2005.
- [2] K. Kunze and P. Lukowicz, "Using acceleration signatures from everyday activities for on-body device location," pp. 115–116, 2007.
- [3] —, "Sensor placement variations in wearable activity recognition," *IEEE Pervasive Computing*, vol. 13, no. 4, pp. 32–41, 2014.
- [4] N. Amini, M. Sarrafzadeh, A. Vahdatpour, and W. Xu, "Accelerometer-based on-body sensor localization for health and medical monitoring applications," *PUC*, vol. 7, no. 6, pp. 746–760, 2011.
- [5] D. Weenk, B.-J. F. Van Beijnum, C. T. Baten, H. J. Hermens, and P. H. Veltink, "Automatic identification of inertial sensor placement on human body segments during walking," *Journal of neuroengineering and rehabilitation*, vol. 10, no. 1, p. 31, 2013.
- [6] T. Zimmermann, B. Taetz, and G. Bleser, "IMU-to-segment assignment and orientation alignment for the lower body using deep learning," *Sensors*, vol. 18, no. 1, p. 302, 2018.
- [7] A. Mannini, A. M. Sabatini, and S. S. Intille, "Accelerometry-based recognition of the placement sites of a wearable sensor," *Pervasive and mobile computing*, vol. 21, pp. 62–74, 2015.
- [8] Y. Maki, S. Kagami, and K. Hashimoto, "Accelerometer detection in a camera view based on feature point tracking," in *IEEE/SICE International Symposium on System Integration*, 2010, pp. 448–453.
- [9] S. Stein and S. J. McKenna, "Accelerometer localization in the view of a stationary camera," in *2012 Ninth Conference on Computer and Robot Vision*. IEEE, 2012, pp. 109–116.
- [10] Y. Huang, M. Kaufmann, E. Aksan, M. J. Black, O. Hilliges, and G. Pons-Moll, "Deep inertial pose learning to reconstruct human pose from sparseinertial measurements in real time," *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)*, vol. 37, no. 6, pp. 185:1–185:15, Nov. 2018.
- [11] Z. Jocy, A. Pour Yazdan Panah Kermani, and D. Roggen, "Multimodal fusion of imus and eps body worn sensors for scratch recognition," *Proceedings of Pervasive Health 2020*, 2020.
- [12] J. Kempfle and K. van Laerhoven, "PresentPostures: A wrist and body capture approach for augmenting presentations," in *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, mar 2018.
- [13] H. Kwon, C. Tong, H. Haresamudram, Y. Gao, G. D. Abowd, N. D. Lane, and T. Ploetz, "Imutube: Automatic extraction of virtual on-body accelerometry from video for human activity recognition," *CoRR*, vol. abs/2006.05675, 2020. [Online]. Available: <https://arxiv.org/abs/2006.05675>
- [14] A. Baak, T. Helten, M. Müller, G. Pons-Moll, B. Rosenhahn, and H.-P. Seidel, "Analyzing and evaluating markerless motion tracking using inertial sensors," in *European conference on computer vision*. Springer, 2010, pp. 139–152.
- [15] J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, and R. Moore, "Real-time human pose recognition in parts from single depth images," *Commun. ACM*, vol. 56, no. 1, pp. 116–124, Jan. 2013.