Low Cost Embedded Multimodal Opto-Inertial Human Motion Tracking System

Mariusz P. Wilk*, Michael Walsh, Brendan O'Flynn
Tyndall National Institute
University College Cork
Cork, Ireland
mariusz.wilk@tyndall.ie* (corresponding author), michael.walsh@tyndall.ie, brendan.oflynn@tyndall.ie

***Abstract*—Human motion tracking systems are widely used in various application spaces, such as motion capture, rehabilitation, or sports. There exists a number of such systems in the State-Of-The-Art (SOA) that vary in price, complexity, accuracy and the target applications. With the continued advances in system integration and miniaturization, wearable motion trackers gain in popularity in the research community. The opto-inertial trackers with multimodal sensor fusion algorithms are some of the common approaches found in SOA. However, these trackers tend to be expensive and have high computational requirements. In this work, we present a prototype version of our opto-inertial, motion tracking system that offers a low-cost alternative. The 3D position and orientation are determined by fusing optical and inertial sensor data together with knowledge about two external reference points using a purpose-designed data fusion algorithm. An experimental validation was carried out on one of the use cases that this system is intended for, i.e. barbell squat in strength training. The results showed that the total RMSE in position and orientation was 32.8 mm and 0.89 degree, respectively. It operated in real-time at 20 frames per second.**

Keywords—Internet of Things, Opto-inertial, Motion, Tracking, Embedded, Low-Cost, Sensor Fusion

I. INTRODUCTION

Multimodal sensor fusion is a common approach in the design of many motion tracking systems. It is based on using more than one sensor modality to measure different aspects of a phenomenon and capture more information about it than what would be available otherwise from a single sensor. Multimodal sensor fusion algorithms often leverage the complementary nature of the different sensor modalities to compensate for their individual shortcomings. This approach is particularly suitable for low-cost and highly miniaturised wearable human motion tracking systems that are expected to perform their function with limited resources at their disposal (energy, processing power, etc.). Such wearable systems can be considered an alternative to the more complex and expensive systems that are commonly used in application spaces such as: motion capture, biomechanical motion analysis or rehabilitation. Some of the most commonly used systems in the industry include the Vicon, OptiTrack or the generally less expensive, but also less robust, Kinect system [1-3]. These systems are classified as "outside-in" systems, i.e. the cameras are outside the tracked object and look in to keep track of points of interest such as Infrared (IR) markers.

In recent years, "inside-out" trackers where processing of position is calculated on the wearable embedded system, have attracted interest of the research community. With the miniaturization and increasing processing power of electronic systems, it is increasingly more feasible to use cameras embedded in the wearable devices themselves as the inside-out trackers, i.e. the camera is mobile and tracks its position based on points of reference in the environment in its Field of View (FoV). This methodology is widely used in mobile robotics to help the robots navigate in buildings using a broad category of algorithms referred to as Simultaneous Localisation And Mapping (SLAM) [4]. However, such inside-out trackers continue to have high processing requirements and require hardware that is found in modern smartphone devices [5]. The addition of vision sensors in wearable motion tracking has many advantages, the most important of which is the ability to determine the absolute position based on some fixed reference points. However, it is a challenging task as cameras impose high processing requirements for real-time operation. The image processing algorithms involved in finding the points of reference require significant computational power. Moreover, the 3-Dimensional (3-D) pose, i.e. its position and orientation, estimation algorithms are generally mathematically complex. In most cases, the use of some form of SLAM is involved and specifically the task of solving the Perspective-n-Point problem (PnP) [6]. However, the miniaturised, low-cost and battery-operated wearable devices do not meet such requirements. One of the leading inside-out opto-inertial motion tracker in the SOA is the IS-1500 system developed by Thales Intersense [7]. It is a highly accurate system with typical error within 2 mm. However, it requires high processing power, with externally connected consumer-grade laptop computer and at least four known points of reference. Thus, it is not a suitable solution for

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) and is co-funded under the European Regional Development Fund under Grant Number 13/RC/2077.

many applications that require low-cost trackers. Therefore, alternative low-cost approaches are developed to overcome these limitations. To this end, the processing requirements of these opto-inertial systems are reduced. Firstly, the number of reference points being tracked is reduced, e.g. IR LEDs. In general, the fewer points the camera is to track, the less computationally intensive the point tracking task is. Secondly, the pose estimation algorithms are simplified. For instance, the solution to the PnP problem requires at least three points of reference and complicated mathematical operations [6, 8]. Recent works in the literature show that the number of points of reference can be reduced to two by integrating an Inertial Motion Unit (IMU) in the tracking system and fusing the two sensor modalities (optical and inertial); using the complementary nature of these two sensor modalities. Two examples of such systems were proposed by two groups of researchers, i.e. Li et al. and Maereg et al. [9, 10]. These two systems are outside-in opto-inertial motion trackers that both use external cameras with an IMU installed in the mobile platform. The 3-D pose is determined by fusing data from the two sensor modalities.

The works of Li et al. and Maereg et al. were the inspiration of the work described in this paper. Whereas their works described outside-in systems, this paper describes the embedded prototype of an inside-out, monocular, opto-inertial tracker that also requires only two points of reference. In this paper, an embedded prototype system is presented and experimentally validated. The described prototype uses low-cost hardware and all computations are carried out on the microcontroller unit (MCU) on the wearable device. This is made possible due to the novel data fusion algorithms designed to operate on low-cost, resource-constrained, embedded systems. The system described consists of a wearable smart sensor system, referred to as Wearable Platform (WP), which incorporates the two sensor modalities, i.e. monocular camera (optical) and IMU (motion). The WP operates in conjunction with two optical points of reference embedded in the ambient environment to enable positional tracking in that environment. In addition, a novel multimodal sensor fusion algorithm is utilized which uses the complementary nature of the vision and IMU sensors in conjunction with the two points of reference in the ambient environment. It determines the 3-D pose of the WP in a computationally efficient way. This prototype system is based on our previous work [11, 12].

II. METHODOLOGY

A proof-of-concept prototype of the inside-out motion tracking system was developed. It is a low-cost embedded system that incorporates vision and IMU technologies in an MCU-based wearable device. The WP runs the proposed novel algorithms for 3-D motion tracking with two external reference points placed in the ambient environment. The objective was to test how feasible it was to successfully implement the proposed system in the context of low-cost and resource-constrained conditions. The general architecture of the system, along with its typical workspace, is shown in Fig. 1.

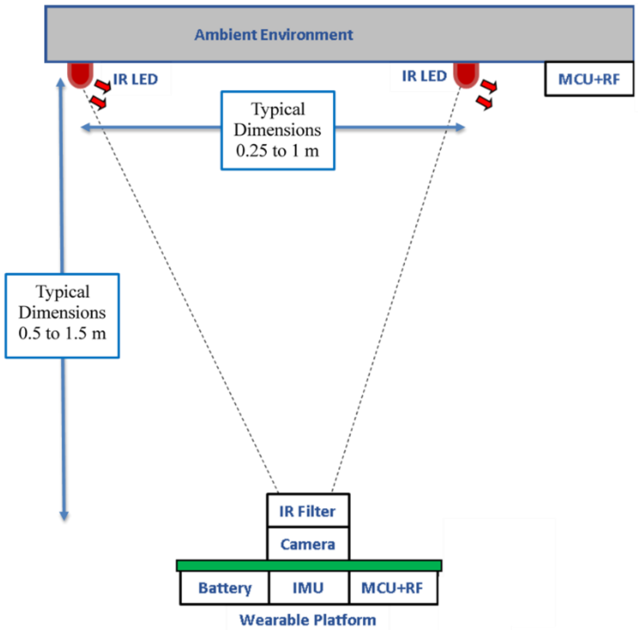
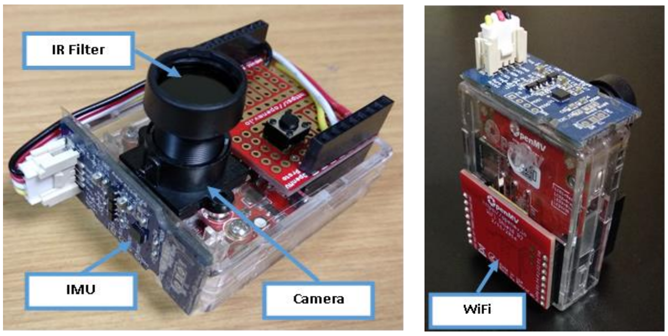


Fig. 1: Wearable Platform with the two IR LEDs and the typical workspace

The prototype implements the proposed system architecture and the novel multimodal sensor fusion algorithm for 3-D pose detection as described in our previous works [11, 12]. The WP incorporates the camera and IMU sensor modalities, along with an MCU and telecommunications capability for controlling the intensity of the two IR LEDs. The typical size of the workspace, or work envelope, is also shown. The size of the work envelope can be varied by changing the distance between the two IR LEDs, their number and positioning; followed by updating the configuration of the embedded sensor fusion algorithm on WP's MCU.

A small form-factor development platform that incorporated a camera module was selected for this task. The OpenMV Cam H7 development board was used in this task which is suitable for wearable applications [13]. This is an MCU-based platform designed for rapid prototyping of projects that incorporate machine vision. It uses an Arm Cortex-M7 STM32H743VI MCU, which is powerful enough to perform real-time image processing tasks while being embeddable in small, low-cost, and energy efficient wearable devices, such as the WP [14]. The WP system includes two sensor modalities, i.e. vision and IMU. To this end, the prototype incorporated the MT9V034 global shutter camera module which comprises of the image sensor and a fixed-focus lens [15]. The camera module includes a custom developed optical IR filter which was attached to the camera's lens, to allow the camera to detect the IR light spectrum only; which is at the same wavelength as that emitted by the IR LEDs [16]. The IMU used for the implementation was the MPU9250 [17]. Additionally, a WiFi shield was added to enable wireless communications [18]. The miniaturised prototype is shown in Fig. 2 (a) and (b); with the associated building blocks of the proposed technology highlighted. The WP works with the two known external points of reference, i.e. the IR LEDs, as shown in Fig. 3.



(a) (b)

Fig. 2: Embedded prototype system: (a) Front view with camera, IMU and IR Filter; (b) Rear view with WiFi shield for communications feature

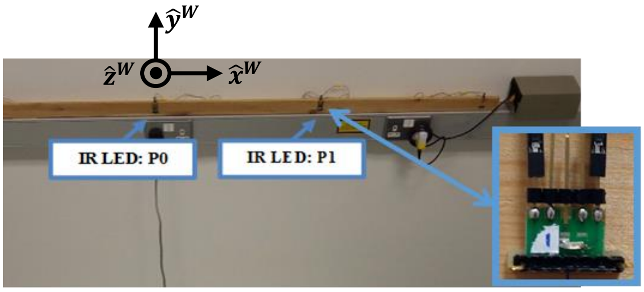


Fig. 3: Two external reference points; IR LEDs; with a 50 cm distance between them; Origin of World Reference Frame is coincident with P0

The WP ran embedded code written in MicroPython which is an implementation of the Python 3 programming language that contains a small subset of standard Python libraries optimised for resource-constrained MCUs [19]. The multimodal sensor fusion algorithm that runs on the embedded WP is described in detail in the following work [12]. This algorithm computes the 3D pose (position and orientation) of the WP as follows. In each frame, the images of the two IR LEDs are captured by the camera and their coordinates on the pixel array are determined using image processing techniques, e.g. point detection. The sensor fusion algorithm uses the geometries formed by the camera and the two points of reference, along with the front-plane pinhole camera model, i.e. pinhole model with image plane in front of the optical centre to compute the 3D pose. A triangle is formed with vertices coincident with the two IR LEDs and the optical centre of the camera. The knowledge of the camera's properties, such as focal length and pixel size, along with the distance between the two IR LEDs enables the use of trigonometry to solve for the lengths of the segments of this triangle, i.e. using similar triangles. The orientation angles of the WP, determined by its IMU, provide the missing pieces of information to compute the complete 3D Pose. Hence, the described algorithm calculates the 3D pose by fusing the input data from visual and inertial sensor modalities. The algorithm was implemented in MicroPython as a custom class that could be imported into the main program's file. Fig. 4 shows a screenshot of the OpenMV Integrated Development Environment (IDE) software during one of the tests of the proposed system. The top-right-hand side of the IDE's window shows the input image frame with the two reference points, i.e. IR LEDs, and their pixel coordinates. The result of the latest 3-D pose computation is also superimposed at the bottom of this input image frame. The 3-D pose was computed based on the input vectors defining the IMU rotation angles and positions of the two points of reference on the image plane *Thzyx* and *P01*, respectively. The *Thzyx* are the Euler rotation angles about z-y-x axes that describe orientation of the WP in World reference frame. The origin of the World reference frame is coincident with IR LED that corresponds to point P0, as shown in Fig. 3. The *P01* is a vector that contains the coordinates of the two IR LEDs on the Image plane, i.e. points' positions in pixel coordinate units. These two vectors are computed at the pre-processing stage that involved the point detection routines in images and the orientation angle estimation step using the IMU data and an open-source sensor fusion algorithm proposed by Madgwick et al. [20]. For each image frame, the pose was computed by making a call to the member function *compute3dPosition* of the *pose3D* object which was an instance of the custom MicroPython class that contained the implemented sensor fusion algorithm. The use of this function is shown in the highlighted line of code in Fig. 4. This line of code corresponds to the Data Fusion block shown in Fig. 5, which describes the general architecture of the embedded algorithm on the WP. The Data Fusion block contains the multimodal sensor fusion algorithm while the other blocks describe the tasks that are carried out in order to condition the input data; prior to passing it to the main Data Fusion block. The input parameters *P01* and *Thzyx*, shown in Fig. 4, correspond to p^I and θ^W in the block diagram, respectively; except for the order of elements in $Thzyx$ which was reversed.

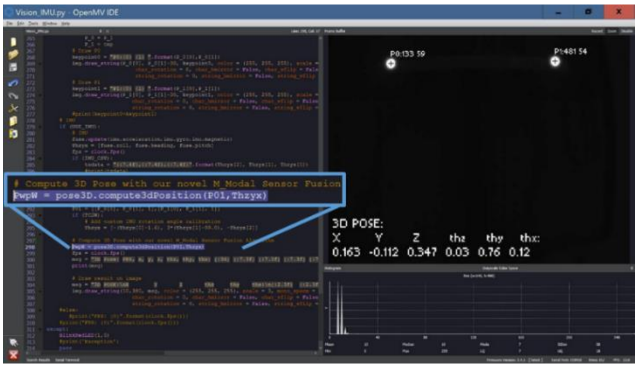


Fig. 4: Embedded prototype Embedded Execution of the Proposed Multimodal Sensor Fusion Algorithm in OpenMV IDE in Real-Time

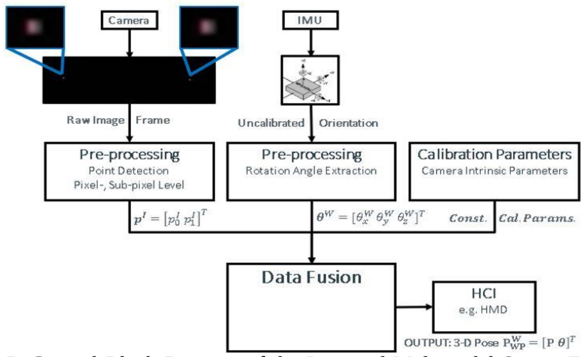


Fig. 5: General Block Diagram of the Proposed Multimodal Sensor Fusion Algorithm, along with the Input Pre-Processing Stages

A. *Performance Evaluation*

The prototype was experimentally evaluated to determine the performance of the embedded version and its feasibility in the context of resource-constrained, MCU-based, wearable applications. The purpose of this experiment was to determine the performance of the embedded prototype system in terms of accuracy and execution time. A specific use case scenario was considered in this study. The repetitive linear motion along a vertical path was examined. It was designed to meet the requirements of one of the considered target application spaces. The specific motion pattern to be simulated was the barbell squat in the context of strength training, hence the repeated linear motion pattern.

B. *Experimental Setup*

The miniaturised wearable prototype system was evaluated in a mobile case scenario that was focused on a repetitive vertical motion pattern. The conditions in this evaluation procedure were such that the camera of the WP was able to capture the two points of references during the experimental runs. The embedded prototype of WP used the MT9V034 camera module [15]. The size of the work envelope was adjusted to ensure that two points of reference were always in the FoV of the MT9V034 camera module. As a result, an appropriate positioning of the embedded WP with respect to the IR LEDs was determined. Specifically, the WP was positioned at $z^W = 1.5\ m$ and the range of motion along the vertical axis in World frame of reference was $y^W \in < -0.55, 015 > m$, while the horizontal position was kept constant at $x^W = 0.25\ m$. The IMU was calibrated and the orientation angle measurements based on its output were used as an input in computing the 3-D pose in real-time. Likewise, the camera on the WP was set to a resolution of 640-by-480 pixels, and calibrated using Zhang's camera calibration procedure [12]. Subsequently, the elements of the intrinsic parameter matrix were used to set the constants in the proposed algorithm by hardcoding them in the MicroPython implementation of this algorithm. These calibration parameters are also shown in the block diagram in Fig. 5. The number of 3-D pose measurements in the experiment was $N > 4700$, while the framerate was more than 20 Frames Per Second (FPS). The 3-D pose was computed in real-time while the WP moved on the rail track in a repetitive manner; over eight up-down cycles. The period of each cycle T was 34 seconds. The WP performed all the functions described in the block diagram in Fig. 5 in real-time during the data acquisition procedure, i.e. while the slider was in motion. The WP and its mounting on the vertical slider are shown in Fig. 6. The camera in the WP tracked two IR LEDs that were mounted on the laboratory's wall, as shown in Figure 3. Additionally, the execution time was recorded to evaluate the processing requirements of the main elements of the embedded software. The three aspects of the embedded software were related to the following: machine vision, IMU data fusion, 3-D pose estimation algorithm. To that end, the IMU, vision and 3-D pose estimation tasks were assessed separately.

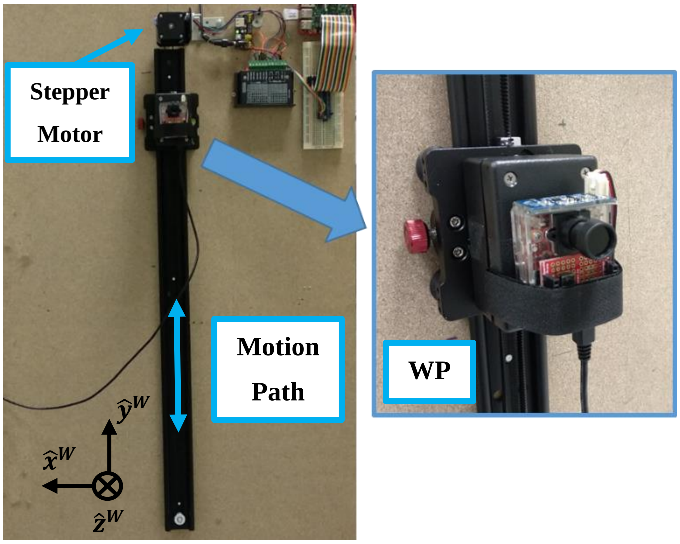


Fig. 6: Experimental Setup: Small-Form-Factor WP was Attached to the Vertical Rail Track Slider; Facing the IR LEDs and the World Frame

The experiments were repeated five times to ensure the results were consistent and statistically significant. The results of one such run were reported in Section III.

III. Results and Discussion

The validation trials have shown that the accuracy of the embedded prototype was consistent and in line with the expectations based on the results of the previous work. The key metric under scrutiny was the RMSE in position and orientation calculations, i.e. the 3-D pose. In these tests, the RMSE in the position and orientation calculations of the WP was computed and is shown in Table 1 and Table 2, respectively. The results show that the overall RMSE in position and orientation was 3.28 cm and 0.8921 degree, respectively. The RMSE in orientation was determined in this experiment using the algorithm proposed by Madgwick et al. [20]. The results show that the RMSE in orientation was generally within the expected range as outlined by Madgwick et al., except for θ_x^W; which was 1.4322 degrees. The relatively high total RMSE in position calculation had several contributing factors. Firstly, the WP was positioned at $z^W = 1.5\ m$, which was at the furthest point along the $\hat{z}^W$ axis in the work envelope for this specific system configuration. Secondly, the 3-D pose was computed in real-time which resulted in the necessity to facilitate the maximum achievable framerate to capture the motion of the WP. To this end, the input images were not corrected for camera lens distortions. It reduced the computational requirements associated with processing the input frames, but it introduced additional error in the coordinates of the reference points. Moreover, the relatively high RMSE in the rotation angle about $\hat{x}^W$-axis θ_x^W, shown in Table 2, adversely affected the accuracy in position calculation along the $\hat{y}^W$-axis in World frame P_y^W. As a result, the RMSE in P_y^W was high which significantly increased the total RMSE in position computation; reaching 5.45 cm. The remaining two position elements of the 3-D pose vector P^W were low. The execution time of the three key aspects of the embedded software was determined. The breakdown of the results is shown in Table 3. The point detection in machine vision was the most time-consuming task. It is understandable since the algorithm needs to look for the points in all pixels in the given input image frame. A further reduction of the camera resolution and/or addition of the temporal element to the point tracking algorithm could mitigate this problem. The algorithm could be programmed to look for the points only in the areas of the image where they are expected to be, based on previous frames decrease the execution time of this task.

Table 1: Experimental validation: RMSE in position calculation

Position	RMSE [m]
$P_x^W[m]$	0.0136
$P_y^W[m]$	0.0545
$P_z^W[m]$	0.0080
Total [m]	0.0328

Table 2: Experimental validation: RMSE in orientation calculation

Orientation	RMSE [m]
$\theta_x^W[deg]$	1.4322
$\theta_y^W[deg]$	0.4444
$\theta_z^W[deg]$	0.3719
Total [deg]	0.8921

Table 3: Experimental validation: RMSE in orientation calculation

Task	Execution Time [ms]
IMU: Orientation Computation	1.3
Vision: Point Detection	30.3
Sensor Fusion: 3-D Pose Detection	17.5

Fig. 7 shows the output of the proposed embedded data fusion algorithm as a function of time, i.e. the position elements of the pose vector in World reference frame P^W over eight up-down cycles of the WP. It shows the accuracy and repeatability of the algorithm in computing the 3-D position over time.

IV. CONCLUSIONS AND FUTURE WORK

This work describes an embedded proof-of-concept motion tracking system. This system was developed for experimental testing of its potential in real-world use cases. It used the MCU-based machine vision development board OpenMV H7 and the MPU9250 IMU. Validation trials have been carried out to measure the accuracy of this wearable solution. The experiments simulated a mobile scenario in the considered sports application space using the example of a barbell squat in a strength training routine. The use case in which the WP is attached to an athlete's back during executing the barbell squat was simulated to track the 3-D motion. Although the experiments were focused only on a single use case, the findings revealed its potential as a low-cost wearable alternative for other similar human various motion tracking application spaces that reach beyond sports such as those in the Industry 4.0.

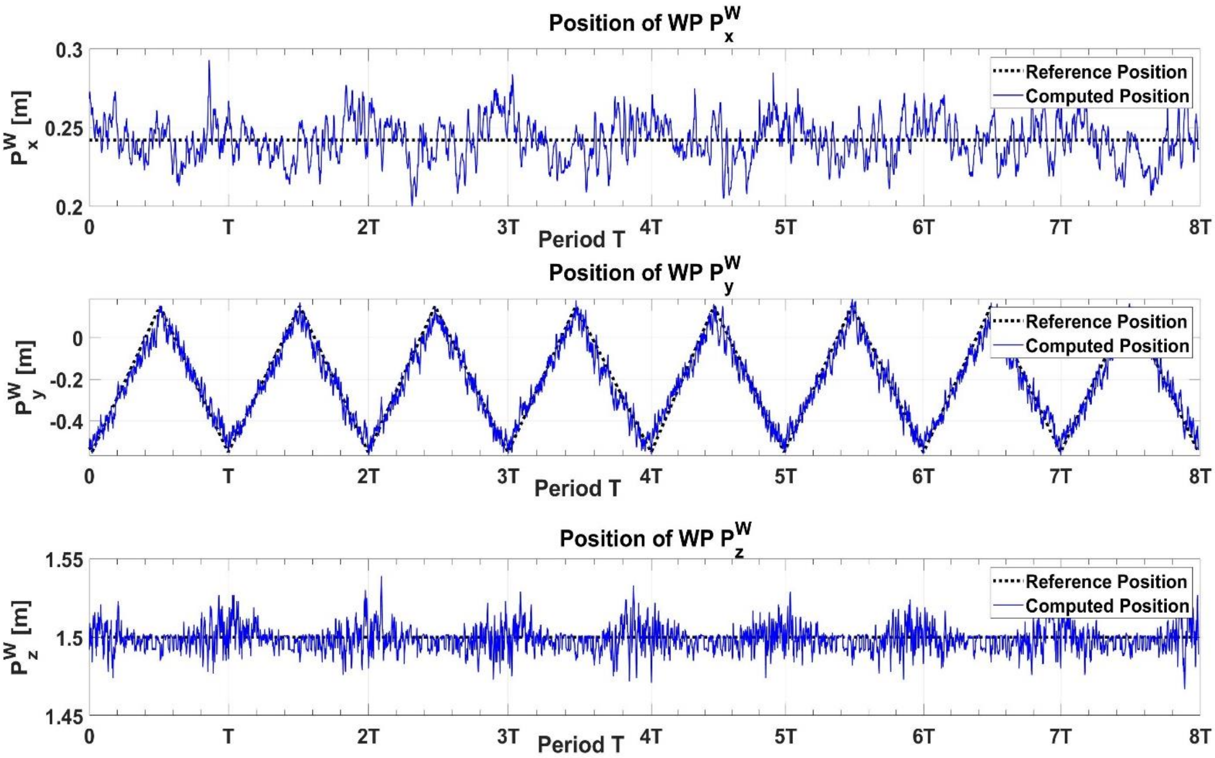


Fig. 7: Experimentally Determined Position of the Small-Form-Factor Version of WP in Repetitive Linear Motion along $\hat{y}^W -$ axis

The accuracy of the embedded version of the WP was consistent with the expectations which were based on our previous simulations and experimental work. Despite using different hardware components and configuration of the embedded WP, the embedded sensor fusion algorithm computed the 3-D pose with a comparable accuracy. Moreover, the embedded WP performed the motion tracking function in real-time, i.e. it ran the image processing/IMU orientation calculation, and the multimodal sensor fusion for the 3-D pose detection algorithms during the data acquisition process. The image processing involved in point tracking was the most resource intensive and time-consuming task in the embedded software. Despite that, it computed the 3D pose at a rate of over 20 FPS. The total RMSE in position and orientation estimation was 3.28 cm and 0.89 degree, respectively. The RMSE in position calculation along the the $\hat{y}^W$-axis in World frame P_y^W was the highest and reached 5.45 cm. It was this high due the lens distortions that were significant near the inflection points along the $\hat{y}^W$ -axis. Nevertheless, the proposed system demonstrates a potential to be a low-cost, inside-out, motion tracking alternative to the existing systems found in the SOA. The results show that it can provide a viable low-cost human motion tracking capability for sports applications, such as the barbell squats in the strength training; to name but a few. A further development work can decrease the RMSE by limiting the main sources of error, i.e. inaccuracies related to camera lens distortions and the IMU. Therefore, an improvement in these two aspects of the system can significantly reduce the error in 3-D pose estimation, thus further increasing the viability of this system in real-world use cases. Furthermore, future tests can be carried out using a battery powered WP.

V. References

[1] Vicon. "Vicon Motion Capture Systems." https://www.vicon.com/ (accessed 20.05.2019.

[2] OptiTrack. "Motion Capture Systems." https://optitrack.com/ (accessed 20.05.2019.

[3] Microsoft. "Kinect V2 Features." Microsoft. https://msdn.microsoft.com/en-us/library/dn782025.aspx (accessed.

[4] E. Guerra, Y. Bolea, A. Grau, and R. Munguía, "Human-robot SLAM in industrial environments," in *2015 IEEE 13th International Conference on Industrial Informatics (INDIN)*, 22-24 July 2015 2015, pp. 390-395, doi: 10.1109/INDIN.2015.7281766.

[5] T. Schöps, J. Engel, and D. Cremers, "Semi-dense visual odometry for AR on a smartphone," in *2014 IEEE international symposium on mixed and augmented reality (ISMAR)*, 2014: IEEE, pp. 145-150.

[6] S. Q. Li, C. Xu, and M. Xie, "A Robust O(n) Solution to the Perspective-n-Point Problem," *Ieee Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1444-1450, Jul 2012, doi: 10.1109/tpami.2012.41.

[7] Intersense. "IS-1500." http://www.intersense.com/pages/70/255 (accessed 25.03.2019, 2019).

[8] V. Lepetit, F. Moreno-Noguer, and P. Fua, "Epnp: An accurate o (n) solution to the pnp problem," *International journal of computer vision*, vol. 81, no. 2, p. 155, 2009.

[9] J. Li, J. A. Besada, A. M. Bernardos, P. Tarrío, and J. R. Casar, "A novel system for object pose estimation using fused vision and inertial data," *Information Fusion*, vol. 33, pp. 15-28, 2017/01/01/ 2017, doi: http://dx.doi.org/10.1016/j.inffus.2016.04.006.

[10] A. Maereg, E. Secco, T. Agidew, D. Reid, and A. Nagar, "A Low-Cost, Wearable Opto-Inertial 6-DOF Hand Pose Tracking System for VR," *Technologies*, vol. 5, no. 3, p. 49, 2017.

[11] M. P. Wilk and B. O'Flynn, "Miniaturized Low-Power Wearable System for Human Motion Tacking Incorporating Monocular Camera and Inertial Sensor Data Fusion for Health Applications," presented at the Smart Systems Integration 2019, Barcelona, Spain, 2019.

[12] M. P. Wilk, M. Walsh, and B. O'Flynn, "Multimodal Sensor Fusion for Low-Power Miniaturised Wearable Human Motion Tracking Systems in Sports Applications," *IEEE Sensors (under review)*, 2019.

[13] L. OpenMV. "OpenMV Cam H7." https://openmv.io/products/openmv-cam-h7 (accessed 13.09.2019, 2019).

[14] S. Microelectronics. "STM32H743VI." https://www.st.com/en/microcontrollers-microprocessors/stm32h743vi.html (accessed 13.09.2019, 2019).

[15] OpenMV. "Global Shutter Camera Module." https://openmv.io/products/global-shutter-camera-module (accessed 6.05.2019, 2019).

[16] E. O. Inc. "1" Diameter, Optical Cast Plastic IR Longpass Filter." Edmund Optics Inc. https://www.edmundoptics.com/optics/optical-filters/longpass-edge-filters/1quot-diameter-optical-cast-plastic-ir-longpass-filter/ (accessed 19.02.2018, 2018).

[17] T. InvenSense. "MPU9250." https://www.invensense.com/products/motion-tracking/9-axis/mpu-9250/ (accessed 29.04.2019, 2019).

[18] O. LLC. "Wifi Shield." https://openmv.io/collections/shields/products/wifi-shield (accessed 13.09.2019, 2019).

[19] MicroPython. "MicroPython." https://micropython.org/ (accessed 14.09.2019, 2019).

[20] S. O. H. Madgwick, A. J. L. Harrison, and R. Vaidyanathan, "Estimation of IMU and MARG orientation using a gradient descent algorithm," in *2011 IEEE International Conference on Rehabilitation Robotics*, June 29 2011-July 1 2011 2011, pp. 1-7, doi: 10.1109/ICORR.2011.5975346.