Ref. Ares(2023)5902231 - 30/08/2023

humantech

D4.1 – Body sensor network with integrated camera approach

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement n° 101058236. This document reflects only the author's view, and the EU Commission is not responsible for any use that may be made of the information it contains.

Project Title	Human-Centred Technologies for a Safer and Greener European Construction Industry.
Project Acronym	HumanTech
Grant Agreement No	101058236
Instrument	Research & Innovation Action
Topic	HORIZON-CL4-2021-TWIN-TRANSITION-01-12
Start Date of Project	June 1, 2022
Duration of Project	36 months

Name of the Deliverable	Body sensor network with integrated camera approach
Number of the Deliverable	D4.1 (D16)
Related WP Number and Name	WP4 – Wearable Technologies for Construction
Related Task Number and Name	T4.1 Inertial sensors - wearable camera integration
Deliverable Dissemination Level	Public
Deliverable Due Date	31. August 2023
Deliverable Submission Date	August 2023
Task Leader/Main Author	Sci-Track (SCT) / Markus Miezal
Contributing Partners	SCT, DFKI, RICOH
Reviewer(s)	Bruno Mirbach, Jason Rambach

Keywords

Inertial sensor network, Visual inertial sensor fusion,

Revisions

Version	Submission date	Comments	Author
v0.1			
v0.2			
...			
...			
...			

Disclaimer

Acronyms and definitions

Acronym	Meaning
EKF	Extended Kalman Filter
IMU	Inertial measurement unit
DoF	Degree of Freedom
HMD	Head Mounted Display
FOV	Field of view
CCS	Camera coordinate system
ICS	IMU coordinate system

Abstract

This document describes the visual inertial sensor network which the workers wear in some use cases within the HumanTech project. The focus of this deliverable lies on the hardware properties as well as its calibration. As part of this hardware, the local processing device will be introduced as well. It is planned, that this powerful device will also host other applications in the context of the worker, such as the localization of Task 4.3 and the exoskeleton controller with intention prediction of Task 4.2 and will exchange data with them. The output of D4.1 is mainly used to predict the wearers intention in T4.2 but it may also interact with T4.3. It will be demonstrated within Pilot 1.

The camera is developed and described by RICOH, while the sensor network was developed by DFKI and SCT together. The integration is performed and described by

The HumanTech project

The European construction industry faces three major challenges: increase the safety and wellbeing of its workforce, improve its productivity, and become greener, making efficient use of resources.

To address these challenges, HumanTech proposes to develop **human-centred cutting-edge technologies** such as wearables for workers' safety and support and robots that can harmoniously coexist with human workers while contributing to the ecological transition of the sector.

HumanTech aims to achieve major advances in cutting-edge technologies that will enable a safe, rewarding, and digital work environment for a new generation of highly skilled construction workers and engineers.

These advances will include:

- **Robotic devices equipped with vision and intelligence** that allow them to navigate autonomously and safely in highly unstructured environments, collaborate with humans and dynamically update a semantic digital twin of the construction site in which they are.
- **Smart, unobtrusive workers protection and support equipment**. From exoskeletons activated by body sensors for posture and strain to wearable cameras and XR glasses that provide real-time workers' location and guidance for them to perform their tasks efficiently and accurately.
- An entirely new breed of **Dynamic Semantic Digital Twins (DSDTs) of construction sites** that simulate in detail the current state of a construction site at the geometric and semantic level, based on an extended Building Information Modelling (BIM) formulation that contains all relevant structural and semantic dimensions (BIMxD). BIMxDs will act as a common reference for all human workers, engineers, and autonomous machines.

The **HumanTech consortium** is formed by 22 organisations — leading research institutes and universities, innovative hi-tech SMEs, and large enterprises, construction groups and a construction SME representative — from 10 countries, bringing expertise in 11 different disciplines. The consortium is led by the German Research Center for Artificial Intelligence's Augmented Vision department.

Contents

1. Introduction

The purpose of this deliverable is to introduce the wearable body sensor network and camera integration for the HumanTech (HT) project. Its goal is to monitor the worker's posture and his position for e.g., intention prediction and as input for an exoskeleton controller. Another part is the hardware platform, which the worker carries on-body and processes the data of the body-worn sensors. It will potentially host the software from multiple partners along with the tracking, e.g., the exoskeleton controller or the camera-based localization.

This deliverable describes the chosen components including their interaction on hardware level, their intrinsic and extrinsic calibration, and which actions have been taken to ensure that subsequent tasks will receive the required data while allowing flexibility in the further cause of the project. Figure 1 show the currently planned data flow. The sensor fusion module receives the data from the sensors and may distribute it between local services and potentially to external endpoints (e.g. a server). Task 4.3, the visual localization may also interact with the sensor fusion module. Based on the received image, it can provide an initial localization and continuously provide a refined pose. It may also use other kinds of data that may be extracted in the sensor fusion module, e.g., a sparse point cloud, or relative poses, which can act as prior for the localization. The exoskeleton module predicts the intention of the wearer based on raw or fused sensor data. All modules will have the opportunity to use wireless network for communication, e.g., to interact with the XR-glasses from Task 4.4.

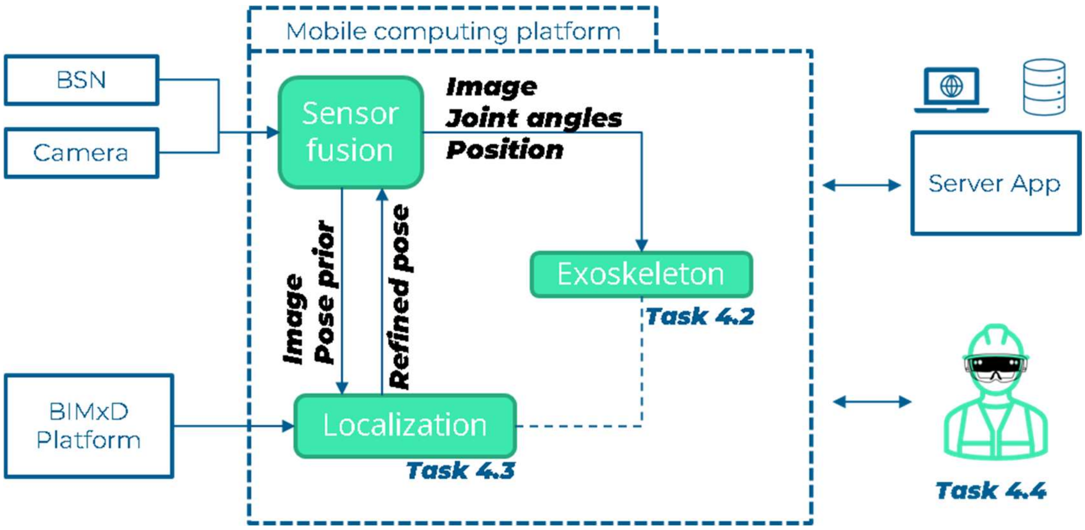

Figure 1: Planned dataflow within WP4 and interaction with other components.

The final prototype should be unobtrusively integrated into the clothing of the worker.

1.1. Basic Concepts

The body sensor network consists of multiple inertial measurement units (IMUs) each of which provide 3D measurements of acceleration (including gravity) through the accelerometer, rotational velocity through the gyroscope and magnetic field through the magnetometer. A typical application is orientation estimation. Exploiting the fact that gravity is captured by the accelerometer, the pitch and roll angle can be determined. Information about yaw is provided by the magnetometer. The latter sensor, however, is easily disturbed, e.g., in proximity of ferro-magnetic materials, current carrying wires or electric motors. These items are particularly found on construction sites, which poses a big challenge for this project. Since each major limb segment carries a sensor unit and magnetic disturbances are a rather local phenomenon, the problem is even harder.

1.1.1. Drift omission by visual observation

In literature, many approaches to overcome magnetic disturbances have been evaluated. Some try to detect or estimate and omit or correct the magnetic disturbance [1]–[3] But for many cases, one may construct a scenario where the approaches may fail, so many publications try to leave out magnetic information completely ([4]–[6]), therefore scarifying any extrinsic (global) yaw information. This approach works for chains of sensors and maintains intrinsic yaw accuracy, if enough motion in certain degrees of freedom (DoFs) is present. Consequently, these approaches fail in static conditions: if no acceleration is present, only gravity can be measured, therefore no yaw information is available and the estimate if forced to follow a noisy gyroscope reading which leads to drifts. In such situations the yaw direction becomes not observable. At this point, optical information can help significantly.

Intrinsic and extrinsic camera images have been used to successfully detect humans and their posture. In most cases a full 3D pose may be obtained [7], however, it suffers from occlusions, depends on proper lighting conditions and has issues with internal rotations of segments which is sometimes barely visible in the images. In contrast to the segment orientations, the joint centers can be reliably detected (apart from occlusions). When a camera is placed on a known position on the body, drift for segments, of which the joint points which are not aligned with gravity can be removed.

1.1.2. Biomechanical calibration

Another aspect where the visual detection of joint points helps is the calibration of the biomechanical model. The basis of the tracking is kinematic model which is driven by a sensor fusion method incorporating the measurements. Typically, this model assumes segments of fixed length which are obtained from anthropometric tables. Besides the fact, that these are not accurate, also the sensors, which are placed on each segment are assumed to be at a certain position and orientation relative to the segment's origin. The transformation at which the sensor is attached on the corresponding segment is called I2S calibration. By observing the joint points in the camera and the reprojection of the current biomechanical model into the camera image there might be the possibility to extract information about the calibration.

1.1.3. Translational drift reduction

A third scenario where a camera can be very helpful is translational tracking or localization. In a pure inertial system, already linear velocity in not observable, i.e., it is impossible to determine the current velocity given the inertial measurements. Usually while tracking, it is assumed that a static condition per segment or ground contact can be detected, and then the drift is bound by zero-velocity updates. While these detections can be erroneous, too early or too late, they potentially lead to positional drift which adds up to the directional drift that occurs when magnetic measurements are dropped.

A camera in turn can determine its position velocity relative to its surroundings (up to scale) based on visual features. So, also in this case, the sensor modalities are complementary.

1.2. Requirements

Hardware-wise there should be a sensor network that is ready to be unobtrusively integrated into clothing. The camera has to be mounted on the body providing a good visibility of the worker's body and the surroundings while it does not hinder the wearer in any way. Ideally, the system should be comfortably wearable for the duration of a shift.

Considering the fusion approach, there are multiple sensors which have to be temporally and spatially aligned in order to produce a singular estimate. Within inertial sensor networks it is already common to perform synchronization and for this project a camera will be used, which comes with the feature of hardware synchronization. This takes away one larger source of error, since delays may cause the estimations to contradict each other. However, in order to allow more flexibility and usage of camera models which are neither equipped with an IMU ex-factory or of industry-grade, we will target asynchronous operation which has also been presented in literature [8], [9].

2. Hardware

2.1. Inertial sensor network

As inertial sensor network, a version of DFKI's sensor network developed in the BIONIC project[1] was chosen. It has already been developed in the industrial context and directly meets some requirements. The sensors are very small, lightweight and connected via textile cables which eases the unobtrusive integration into clothing and can be comfortably worn.

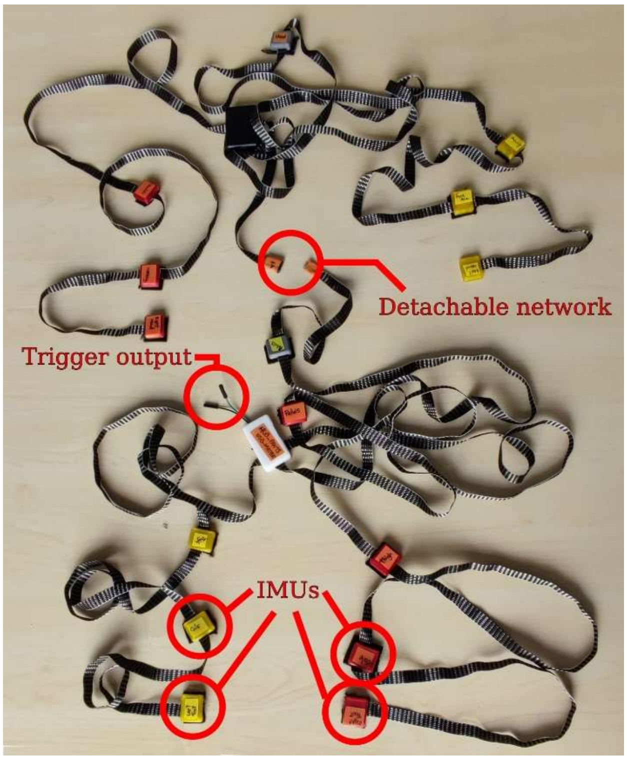

Figure 2: Sensor network with textile cables and features.

It is completely wired, which on one hand ensures continuous and privacy protected communication of sensor data and on the other hand it prevents data loss due to wireless communication, especially when many participants share the same communication technology. Furthermore, the network is powered by the device which

[1] https://cordis.europa.eu/project/id/826304

reads the data (cf. Section 2.4) which in turn is powered by an external battery pack, so that the runtime of the software scales with this battery's capacity.

This modularity of its components extends to the network itself. It can be split apart into the lower and upper body. Initially, the whole body will be tracked, since localization is a topic and especially the sensors on the lower body carry information about locomotion, however, it is possible that the camera's localization is precise and frequent enough to make the lower body obsolete. In this case, it may be detached.

The extension done in HumanTech is the introduction of the camera imu unit which is created by placing the head IMU on the helmet to ensure rigid connection to the camera. Potentially, the connection to the HMD is rigid as well, which can then serve as basis for synchronizing to and localizing the VR Headset.

To synchronize the camera, the sensor network was extended with a trigger output which serves a standard TTL-trigger signal.

2.2. Backup solution

Since only one version of the sensor network exists, for research and development a XSens Awinda system[2] will be used as an alternative. Each sensor is wireless with its own battery, which can keep the sensor in operation for about 6 hours. The data is captured on a base station which is connected to a PC and also provides a trigger output. To transfer the trigger signal to the camera, a SyncBox and an output device have been developed.

Figure 3: SyncBox (left) which reproduces an input signal on the remote trigger device (right).

[2] https://www.movella.com/products/wearables/xsens-mtw-awinda

The SyncBox gets the input trigger from the Awinda base station and adapts its own trigger signal to match this input. The trigger is then wirelessly transferred to the output device, where it is reproduced in the same way, then the original sensor network.

2.3. Camera

Our new camera is designed to achieve three goals, compactness which doesn't hinder the wearer, being synchronous with other devices to solve localization drift caused by magnetic interference to IMUs and wide field of view to capture the wearer's body and environment at once. Traditional fish-eye lens-integrated cameras have been bisected into two kinds of products, consumer models and development models. Consumer-oriented cameras are very compact and easy to take images and videos. However, they don't have functions such as hardware trigger and low-latency image transmission to other devices, which disrupts time-synchronized hardware integration. On the other hand, the cameras aimed to be implemented into other devices have various kinds of functions including hardware trigger and low latency of image transmission. However, they are bulky and uncomfortable to be used as wearable devices. Therefore, it has been necessary to develop a new camera satisfying the advantages of both type of camera at once for the three goals. The technical requirements towards the camera are jointly discussed and defined between sci-track and RICOH. The camera consists of two fish-eye optical components for capturing spatial information with wide field of view and two sets of RGB image sensor and processing unit for converting the captured images and transferring them to on-body computing platform with lower latency.

Table 1 shows the specification of images sensors and processing units provided by IDS imaging GmbH. Thanks to the optical design including multiple deflection of light path inside the component, the component is small compared to the traditional fish-eye lenses for 1-inch image sensor manufactured by FUJIFILM Corporation, which contributes to comfortable usage of the camera attached onto the wearer's body or equipment without disturbing their motion. As shown in figure 5, the volume of our camera is smaller than the combination of pre-existing industrial camera and fish-eye lens by 9%. The image sensor exploited for this camera is PYTHON 5000 and manufactured by ON Semiconductor Corporation. Its size is 1 inch, and the resolution is 2592 by 2048 pixels. The larger image sensor and its pixel enable operation in the dark environment, such as construction sites surrounded by construction equipment and cover sheets.

Table 1: Specification of camera

Unit	item	Specification
Sensor unit	Image sensor	PYTHON 5000
	Sensor size	1 inch
	Resolution	2592 × 2048 px
	Maximum framerate	73.0 fps
	Shutter type	Global
Lens unit	Field of View	> 180 deg.
IMU	Model	RH40127RH

The sensor unit has three kinds of acquisition mode, free-run, software and hardware trigger. On free-run the images are acquired at the frame rate defined by software. Software and hardware trigger are activated by software commands or electrical signal with the level of 5V to a GPIO pin, respectively. For the hardware trigger, the latency is relatively shorter than software trigger, which is at microsecond order and also negligibly small compared to the sampling rate of IMUs. There are various kinds of controllable settings of image acquisition such as frame rate, brightness, exposure time, digital gain and binning factor. Provided that the use-case and its environment drastically vary, the flexibility of image settings is necessary to constantly provide images at stable quality.

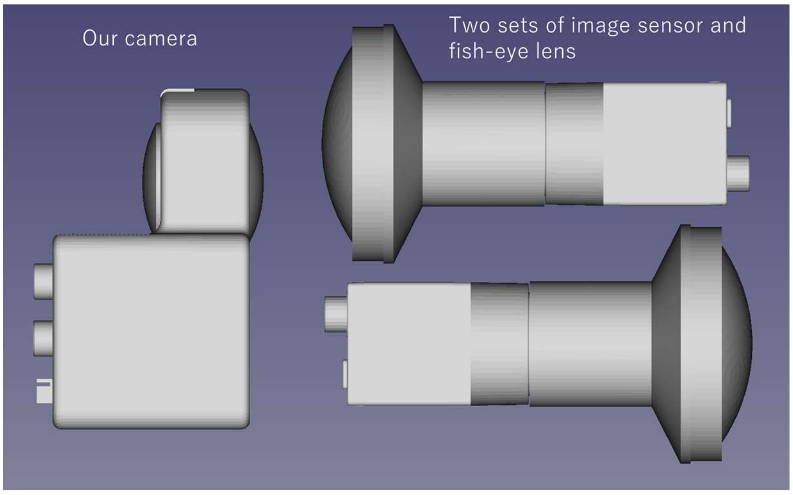

Figure 5: Size and shape comparison between our camera and combination of 1-inch sensor and fish-eye lens at side view.

2.3.1. Housing

Optical components and image sensors are integrated into one housing to realize compactness. Figure X shows the exterior view of the camera. Two fish-eye lenses are facing in opposite directions to cover wide space. Especially for one lens, there is no obstacle at the peripheral area to avoid occlusion on the image. On the other side of the camera, there are processing boards for two image sensors, which are partially covering the image of the other fish-eye lens. However, reduced field of view is expected not to affect the function of localization of the wearer. The housing is made of aluminum with the anodizing process on its surface. Aluminum has high heat conductivity of around 220 W/K*m and light density of 2.71 g/cm^3, which is suitable for heat dispersion coming from image sensors and FPGA on the processing units and light-weight wearable scenario. On the bottom of the housing, 1/4 UNC thread is provided for the mounting and fixing camera onto other hardware such as a helmet with screws.

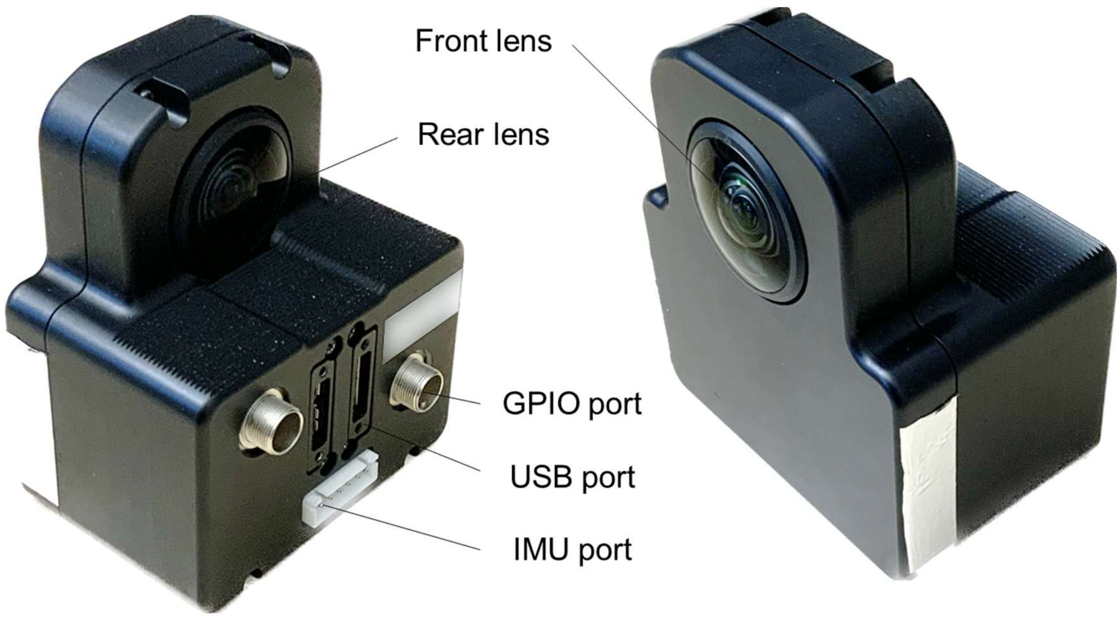

Figure 4: Exterior view of the camera

2.3.2. Mounting

The design of mounting the camera onto a helmet must satisfy several requirements such as proper positioning to capture the wearer's motion without occlusion, weight balance and avoidance of conflicting with other wearable devices. Figure 4 shows the exterior view of the camera mounting. The camera is connected and fixed to HoloLens 2 or Trimble XR10 by 3d-printed objects and screws. There are two mounting positions, central mounting and side mounting. The workers can choose one of the positions depending on the use-case, where they need to put HoloLens 2 or Trimble XR 10. On both positions, the camera is aligned so that the front lens can look down and capture the wearer's body and the rear lens can look up and always record the environmental information for re-localization. Figure 6 (left) shows the captured image with central mouting. The front lens captures both the wearer and environment thanks to the wide field of view. In Figure 6 (right), an example detection from PoseFormerV2 can be seen.

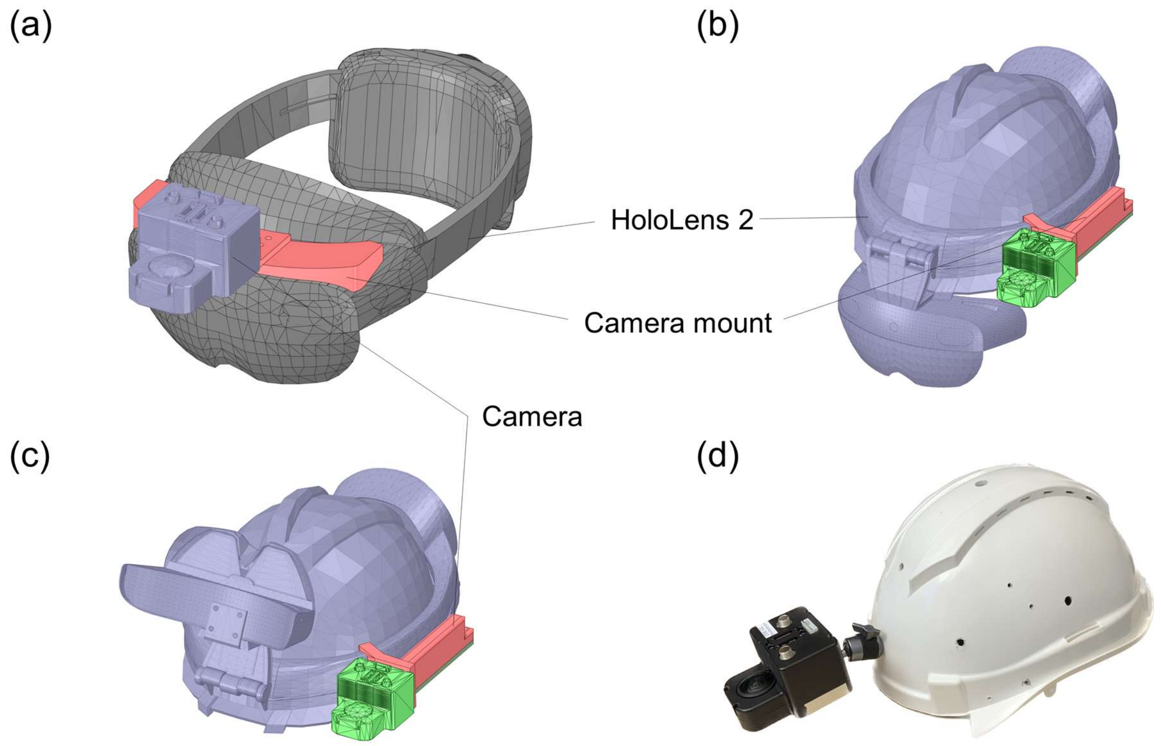

Figure 5: (a) Central mounting of camera on HoloLens 2, side mounting on Trimble XR 10 when HoloLens 2 is (b) under use and (c) flipped up and (d) mounting without other wearable devices.

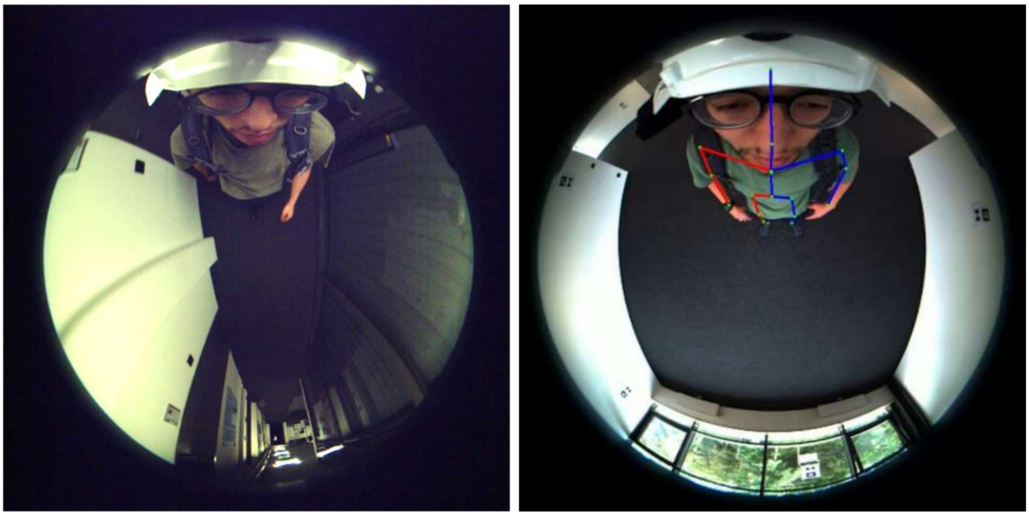

Figure 6: Captured image with front lens (left); Example detection of PoseFormer2[3].

2.4. On-body computing platform

The on-body platform has multiple purposes:

1. It powers the camera and the sensor network,
2. It receives camera and IMU data,
3. It hosts several applications, that process the data and
4. It distributes the data over network, if required.

We chose Qualcomm RB5, as the powerful system to handle these tasks, which already demonstrated its ability in the BIONIC project. It could run a full body tracking at 100Hz in real-time, using 8.08s per sample at a power consumption of 3.44W. During 45 minutes of tracking, the surface temperature of the chip stayed below 40°.

[3] https://github.com/QitaoZhao/PoseFormerV2

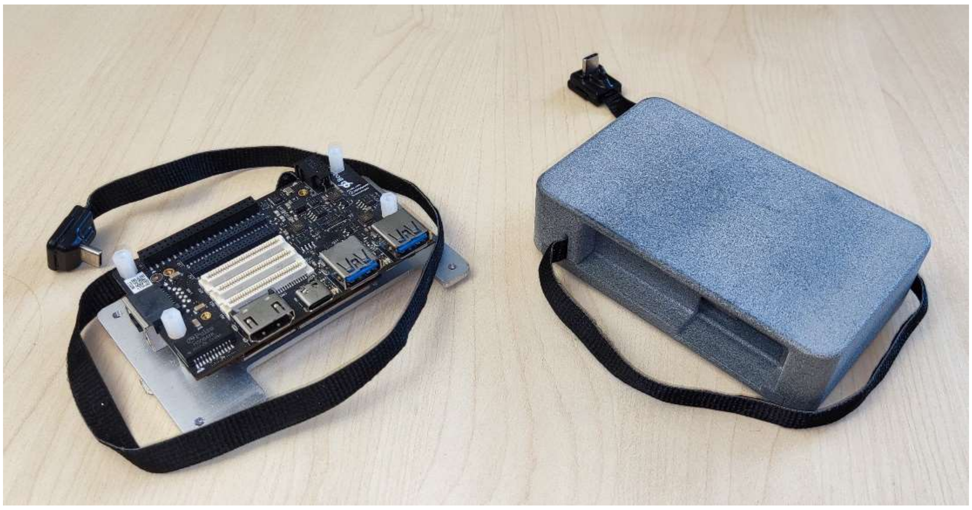

Figure 7: The Qualcomm Robotics Platform RB5 with and without case.

Since the heavy calculations of the body tracking run mostly on a single core, other cores are available to other tasks, such as the localization (Task 4.3) or the exoskeleton controller (Task 4.2). Running them on the same machine as the application that produces their input lowers the communication delay and increases the mobility of the wearer (an external computation device must be reachable for the hardware). For applications which use neural networks, a hexagon DSP is in place to carry the load aside from the CPUs (using TensorFlow Lite).

Keeping the data on the person and not sending it to external machines protects the privacy of this personal data. However, in case the processing power of this device is not sufficient, and the wearer agrees, the device can stream the data over network to an external PC for processing.

2.5. Full setup

Currently, the IMUs network is placed on a Velcro suite. The camera is mounted on the helmet as described above in section 2.3.2, (see Figure 5 (d)). After the setup, work clothing can comfortably worn above (see Figure 8 (left)). Given the current inertial approach, a skeleton estimate can be generated. A picture of the tracking result with a corresponding camera image can be seen in Figure 8 (right).

Figure 8: Setup picture with and without work clothing (left); Sample recording of data at the current level of integration (right).

3. Fusion approaches

3.1. Calibration

3.1.1. IMU calibration (refinement)

The IMU sensors are factory calibrated, however, some parameters may change, in particular the biases. To correct for those, a box calibration is performed. All sensors are somehow rigidly connected to each other, e.g., fixed in a box, or tightly wrapped with straps, then the bundle is turned on all 6 sides, where data is captured over a short interval. The gyroscope biases can be found easily by averaging all samples from the static periods. From the same data, the accelerometer bias can be found, since during the static periods, only gravity has been observed, sampling a sphere with gravity as radius. Consequently, the sphere center is the acceleration bias.

Note that the biases may change over time and with temperature change, so a recalibration before capturing is planned.

3.1.2. Camera calibration

The fisheye cameras have a wide FOV (> 200°) which allows to see points which are behind the camera. This requires a dedicated distortion model. In [10] some of these models are compared. The proposed double sphere model seems to be the best in terms of accuracy and efficiency and will possibly be adopted. Due to availability, however, [11], which has also shown successful calibrations for large FOV cameras, is used for now.

The camera calibration is constant.

3.1.3. CamIMU calibration

Between the IMU and the camera is a spatial and temporal offset. The temporal offset, if not synchronized, must be determined at every run, and may vary over time. In contrast, the spatial offset is constant and is defined by the transformation between the camera coordinate system (CCS) and the IMU coordinate system (ICS). The setup used to calibrate the HumanTech CamIMU unit is based on a gravity aligned chessboard pattern (see Figure 9). The geometry of the chessboard is known, so that the detected 2D points have corresponding 3D points, allowing to reconstruct the camera pose.

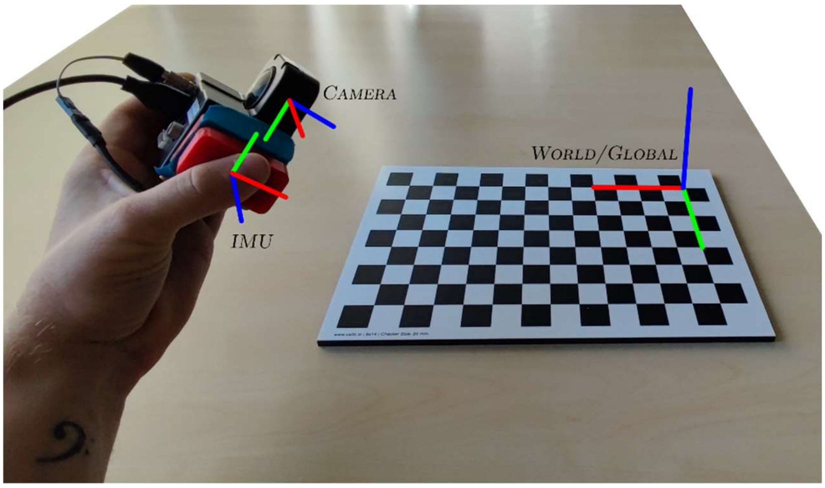

Figure 9: Setup for camIMU calibration: the goal is, to find the relative pose between the camera and the IMU.

A sequence of camera images along with inertial data is recorded, so that it contains some static postures, but also moderate motion which allows the camera to detect the chessboard. The first step is synchronization.

3.2. Synchronization

Since the CamIMU calibration is an offline process, the synchronization is based on the whole dataset. It is known that the sensors are rigidly connected, therefore the magnitude of rotational velocity is the same at every point on this rigid body. The rotational velocity of the camera can be obtained through the chessboard detection which provides a continues 6 DoF camera pose. By differentiating the rotational part by time, the rotational velocity is found. Its norm should be the same as this of the gyroscope, however, shifted in time.

As measurement for how synchronized the values are, the Pearson correlation is used. Given a time offset, the IMU stream is shifted and interpolated at the camera timestamps. Then, all matching pairs of rotational velocity magnitude from both sensors are gathered and the correlation is calculated.

Note, that a gradient based optimization of the time offset will likely fail, since due to the nature of rotational velocity magnitude there is high chance of self-similarity. Instead, a feasible time interval is chosen in which the time offset is located (the sensors are started roughly at the same time so that a few seconds are practical). Based on this, the correlations for all time offsets within the interval with a resolution of twice the IMU frequency, are calculated. The largest one is the sought time offset for calibration.

3.3. Alignment

The recorded calibration sequence, but particularly the static postures are used to determine the relative orientation between the sensors, exploiting the fact, that the chessboard normal is aligned with gravity. In contrast, the relative position requires motion to be estimated. By expressing the sensor position of one sensor in terms of the other sensor and differentiating this twice by time, one can estimate the acceleration of the other sensor. The acceleration of the camera can be derived from its poses and then compared to the accelerometer measurement. Note that gravity has to be added. The relative position is found by minimizing the difference between the estimated and measured acceleration

4. Conclusion

In this document, the visual inertial body sensor network for HumanTech has been presented. An existing inertial body sensor network has been enhanced with a trigger output to drive a newly developed industrial dual fisheye camera which is mounted on the worker's helmet. The data will be captured on a local processing device with a strong CPU and dedicated hardware for neural network computations, allowing to host several applications in the context of the worker. For example, the software of Task 4.3 can run on it, which can use the tracking result as prior and feed refined poses back, thus relating the tracking tt the BIMxD model. The main recipient of the tracking result will be the exoskeleton developed in Task 4.2. For intention prediction and exoskeleton control, all (fused) and raw data can be used. This comprises joint angles, positions (relative to the BIM model), (linear) velocities and accelerations, as well as, rotational velocities and the raw image and sensor data.

The hardware is flexible and can adapt to the project's needs as they evolve. For example, the sensor network can be split into upper and lower body if the output of T4.3. is frequent enough, so that the lower body becomes obsolete. If the original hardware fails, an external sync device was created to allow to use 3rd party hardware (particularly XSens Awinda) as fallback.

The camera hardware is a novel dual fisheye industrial camera with the capability to be electronically triggered. It provides high resolution data with low latency and will be attached to the wearer's head. Together with the head IMU, it forms a visual inertial unit. It will support the inertial tracking to reduce extrinsic and intrinsic heading drift and potentially refine the body model and its I2S calibration.

With the hardware defined and calibrated, an initial fusion approach, based on hardware synchronized data, will be implemented and evaluated. The goal will be an asynchronous fusion.

5. References

[1] G. Ligorio and A. M. Sabatini, "Dealing with Magnetic Disturbances in Human Motion Capture: A Survey of Techniques," *Micromachines*, vol. 7, no. 3, Art. no. 3, 2016.

[2] T. Harada, T. Mori, and T. Sato, "Development of a tiny orientation estimation device to operate under motion and magnetic disturbance," *The International Journal of Robotics Research*, vol. 26, no. 6, Art. no. 6, 2007.

[3] M. Miezal and Weidmann, Alexander, "Bionic: D3.2: Online biomechanical parameter estimation for body parts." European Commission, Apr. 16, 2021. Accessed: Aug. 10, 2023. [Online]. Available: https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5dfb2611e&appId=PPGMS

[4] M. Miezal, "Models, methods and error source investigation for real-time Kalman filter based inertial human body tracking," PhD Thesis, Kaiserslautern University of Technology, Germany, 2021. Accessed: Aug. 10, 2023. [Online]. Available: http://www.dr.hut-verlag.de/978-3-8439-4664-3.html

[5] K. Eckhoff, M. Kok, S. Lucia, and T. Seel, "Sparse Magnetometer-free Inertial Motion Tracking – A Condition for Observability in Double Hinge Joint Systems," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 16023–16030, Jan. 2020, doi: 10.1016/j.ifacol.2020.12.403.

[6] M. Kok, K. Eckhoff, I. Weygers, and T. Seel, "Observability of the relative motion from inertial data in kinematic chains," *Control Engineering Practice*, vol. 125, p. 105206, Aug. 2022, doi: 10.1016/j.conengprac.2022.105206.

[7] D. Tome, P. Peluse, L. Agapito, and H. Badino, "xR-EgoPose: Egocentric 3D Human Pose From an HMD Camera," presented at the Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 7728–7738. Accessed: Aug. 08, 2023. [Online]. Available: https://openaccess.thecvf.com/content_ICCV_2019/html/Tome_xR-EgoPose_Egocentric_3D_Human_Pose_From_an_HMD_Camera_ICCV_2019_paper.html

[8] P. Furgale, T. D. Barfoot, and G. Sibley, "Continuous-time batch estimation using temporal basis functions," in *2012 IEEE International Conference on Robotics and Automation*, St Paul, MN, USA: IEEE, May 2012, pp. 2088–2095. doi: 10.1109/ICRA.2012.6225005.

[9] P. Geneva, K. Eckenhoff, W. Lee, Y. Yang, and G. Huang, “OpenVINS: A Research Platform for Visual-Inertial Estimation,” in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, May 2020, pp. 4666–4672. doi: 10.1109/ICRA40945.2020.9196524.

[10] V. Usenko, N. Demmel, and D. Cremers, “The Double Sphere Camera Model,” in *2018 International Conference on 3D Vision (3DV)*, Sep. 2018, pp. 552–560. doi: 10.1109/3DV.2018.00069.

[11] D. Scaramuzza, A. Martinelli, and R. Siegwart, “A Flexible Technique for Accurate Omnidirectional Camera Calibration and Structure from Motion,” in *Fourth IEEE International Conference on Computer Vision Systems (ICVS’06)*, Jan. 2006, pp. 45–45. doi: 10.1109/ICVS.2006.3.