

# Design Space Exploration on Efficient and Accurate Human Pose Estimation from Sparse IMU-Sensing

Iris Fürst-Walter , Antonio Nappi, Tanja Harbaum , Jürgen Becker 

**Abstract**—Human Pose Estimation (HPE) to assess human motion in sports, rehabilitation or work safety requires accurate sensing without compromising the sensitive underlying personal data. Therefore, local processing is necessary and the limited energy budget in such systems can be addressed by Inertial Measurement Units (IMU) instead of common camera sensing. The central trade-off between accuracy and efficient use of hardware resources is rarely discussed in research. We address this trade-off by a simulative Design Space Exploration (DSE) of a varying quantity and positioning of IMU-sensors. First, we generate IMU-data from a publicly available body model dataset for different sensor configurations and train a deep learning model with this data. Additionally, we propose a combined metric to assess the accuracy-resource trade-off. We used the DSE as a tool to evaluate sensor configurations and identify beneficial ones for a specific use case. Exemplary, for a system with equal importance of accuracy and resources, we identify an optimal sensor configuration of 4 sensors with a mesh error of 6.03 cm, increasing the accuracy by 32.7 % and reducing the hardware effort by two sensors compared to state of the art. Our work can be used to design health applications with well-suited sensor positioning and attention to data privacy and resource-awareness.

## I. INTRODUCTION

Assessment of human motion enriches many applications in sports, rehabilitation or work safety to monitor the movement quality and correctness [1]. These applications require both an accurate Human Pose Estimation (HPE) and sensitive processing of the underlying personal data. Especially in rehabilitation applications, privacy must be carefully respected as the acceptance of continuous monitoring systems depends on data security. To preserve personal information, data locality and local computing is preferred over processing on remote systems. However, due to battery-powered operation, local computing systems have to operate on a strict energy budget, which can be addressed by partitioning the data processing on different compute nodes [2] or data-efficient motion capture (mocap) with Inertial Measurement Units (IMUs) instead of common camera sensing [3].

For IMU-based human motion tracking systems, the number and positioning of the sensors is crucial for an accurate HPE. The more sensors are attached to the human body, the more motion information is captured, but on the other

This work has been supported by the project “Stay young with robots” (JuBot). The JuBot project was made possible by funding from the Carl-Zeiss-Foundation.

The authors are with the Institut fuer Technik der Informationsverarbeitung, Karlsruhe Institute of Technology, Germany. {fuerst, harbaum, becker}@kit.edu

Code available online at <https://github.com/itiv-kit/dse-sparse-imu>.

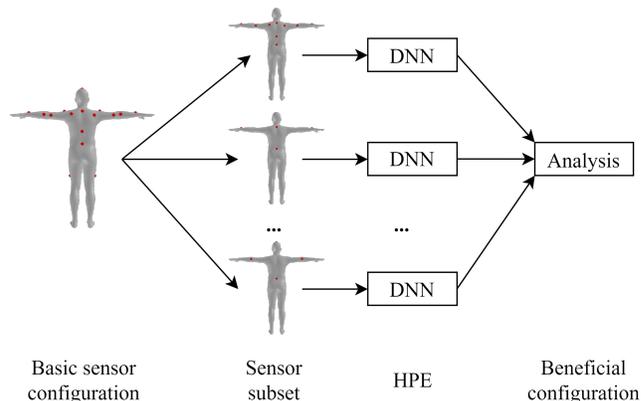


Fig. 1. **Methodology of our Design Space Exploration.** We define a basic sensor configuration and synthesize IMU-data for sensor subsets. With this data, we train a deep neural network and evaluate diverse error metrics for HPE of each subset. Finally, we analyse all experiments to identify a beneficial sensor positioning.

hand, the hardware resource overhead, concerning e.g. energy consumption, available bandwidth or wiring, increases. Therefore, a trade-off between accuracy and resources exists and a compromise for sparse IMU-sensing has to be found in each use case individually. To address this trade-off and establish a general methodology to evaluate systems with different sensor configurations, we perform a Design Space Exploration (DSE) on the number and positioning of IMU-sensors on the human body and propose a combined metric for assessment of the accuracy-resource trade-off. This metric allows to weigh the importance of hardware resources in relation to prediction accuracy defined by the system designer in a specific use case.

Since mounting the sensors is time-intense and error-prone, we synthesize IMU-data for virtual sensors defined on the body model in mocap datasets. To rank different sensor configurations, we train a Deep Neural Network (DNN) on the synthesized IMU-data and compare the resulting HPEs with our combined metric.

We summarize our contribution as

- Synthesis of IMU-data from a body model dataset with additional noise, including ground truth labeling for supervised learning of HPE.
- Automated DSE of variable number and positioning of sensors on the human body with analysis of different sensor setups.
- Evaluation of beneficial sensor configuration in terms of the accuracy-resource trade-off of the HPE using our

introduced combined metric.

Real system design of body-mounted IMU motion tracking is time-consuming, where our DSE can be used as a tool to identify beneficial positions and accelerate the design of fabric-integrated sensor systems. We consider our investigation of sparse IMU-sensing to enable data-efficient and at the same time resource-aware health systems relying on an accurate HPE to pave the way for appropriate application in movement monitoring and pose correction.

## II. RELATED WORK

Human Pose Estimation (HPE) from sparse IMU-sensing and the assessment of different sensor configurations relies on several areas spanning from the model estimating the HPE, used learning data to positioning of the sensors. Previous work in those areas will be presented in the following.

### A. Deep learning based Human Pose Estimation (HPE) from sparse IMU-sensing

Different deep learning based models were investigated to estimate human pose from sparse IMU-data, enabling applications with lower mounting time and higher comfort due to fewer sensors.

**Sparse Inertial Poser (SIP)** [4] showed the potential of HPE from a sparse IMU-configuration of six sensors to enable human motion capture in the wild. The human pose is estimated offline in two stages. The authors further introduced a new evaluation metric, the SIP-error, which only takes the orientation error of upper arms and thighs into consideration. However, this approach is not suited for real-time application.

**Deep Inertial Poser (DIP)** [5] followed SIP with the main goal to improve applicability in online-systems. Real-time capability is enabled by considering a limited number of past and future data and a Recurrent Neural Network (RNN) is used to directly estimate body model parameters from orientation and acceleration inputs. The authors further published a synthesis method to generate IMU-data from mocap datasets. Therefore, motions are mapped to a body model, on which virtual IMU sensors are defined. For each virtual sensor, orientation and acceleration are calculated.

**TransPose** [6] introduced a two-stage approach to estimate pose as well as global translation of human motion. Similar to DIP, bidirectional RNNs are used with Long Short-Term Memory (LSTM) cells to estimate human pose. The authors claim to be more accurate and computationally more efficient and the model achieves processing at 90 fps.

**Physical Inertial Poser (PIP)** [7] was recently published and extends TransPose by a physical-aware model to estimate human pose, initial joint torques and ground reaction forces. The term *physical-aware* refers to the property of respecting physical constraints like jitter and ground penetrations.

Table I gives an overview of the currently available HPE-systems. It can be seen that PIP achieves the most accurate prediction, but it deployed the most complex estimator.

### B. Database

In deep learning, the underlying database is crucial to train a reliable model. Therefore, commonly used datasets for HPE are presented.

**DIP-IMU** [5] recorded solely IMU-data captured by an Xsens motion tracking system with 17 sensor nodes. The placement on the body is similar to Vicon, but not identical. The recording volume is 92 minutes of 330,178 frames at 60 fps. Nine male and one female subject performed five different motions classes, including controlled arm and leg movements, locomotion, jumping jacks or boxing, and interactions with objects while sitting.

**AMASS** [8] is an actively developing collection of mocap datasets, including Skinned Multi-Person Linear Model (SMPL)-data of currently 24 datasets. Overall, 500 subjects performed 17,916 motions, totaling a recording volume of 3,772.45 minutes at 60 fps. Collected motions include whole-body pose motions as well as detailed hand/finger exercises.

In addition to directly recording IMU-data from real movements, [5] proposed synthesis from mocap datasets. Therefore, the authors placed virtual IMU sensors in AMASS and generated orientation and acceleration data for each sensor in each frame of the motion.

**Skinned Multi-Person Linear Model (SMPL)** [9] is a common body model to generically represent human motion and is used in the AMASS dataset. As shown in Figure 2, SMPL simplifies the human body to a skeletal system of 24 segments with 23 joints and provides a wide range of body shapes. The shape consists of a mesh of 6,890 vertices and deforms naturally with a given pose, mimicking soft-tissue dynamics. For the synthesis of IMU-data, a virtual sensor can be placed at any vertex of the SMPL mesh.

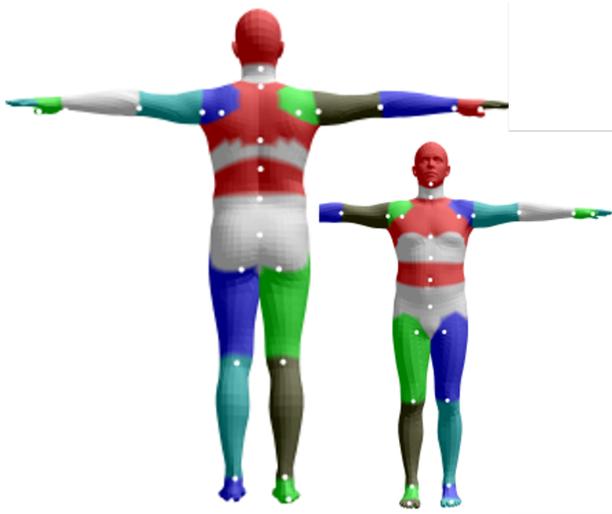
### C. Sensor positioning

**Xsens**<sup>1</sup> is a commercial mocap system with 17 wearable IMUs. It provides a sensor per limb segment, shoulder, head, sternum and pelvis which is considered the root node. This system has been used in, e.g. DIP to collect the DIP-IMU dataset, but the high amount of sensors requires high processing effort.

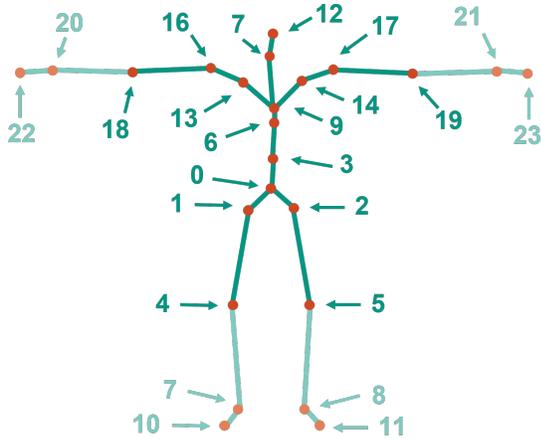
An optimal sensor positioning for instability detection was investigated by Steffan et al. [11]. They evaluated different IMU sensor configurations on the human body, including up to six sensors out of a basic configuration of 34 sensors, which was based on the optical marker placement of Master Motor Map (MMM) [12]. For the DSE, they emulated IMU-data from the Whole-Body Human Motion Database [13], [14] which is included in AMASS. Depending on the F1-score for the instability detection, the best suited sensor configuration and best fitting model were identified. The best instability detection was achieved with six sensors mounted on the right foot, wrists, left elbow, sternum and pelvis.

For motion classification of different displacement motions, Patzer et al. [15] investigated a minimal sensor setup on an exoskeleton. Instead of exhaustively testing every sensor

<sup>1</sup><https://www.xsens.com>



(a) Segmentation. Images from [9]



(b) Skeleton with Joint Indices. Inspired by [10]

Fig. 2. **Segments and joint positions of SMPL** [9]. (a) Segmentation of SMPL, where the white dots depict joints of the body model detailed in (b), shaded joints are not considered in [5] and equally not in ours. (b) corresponds to the back view of the model, i.e., left side of (a)

configuration, they applied a wrapper-based method for three IMUs and seven force sensors. Therefore, they started the DSE with evaluation of a single-sensor-system and defined the sensor with the most accurate motion classification as basic subset. Iteratively, one new sensor was added to the basic setup and the new best subset was identified. As a result, they identified a subset of six sensors to achieve the same accuracy as the basic configuration of ten sensors.

The latter DSEs both identify an optimal sensor configuration for their specific use case of instability detection and motion classification, but lack generalization for other applications and do not consider the hardware resources.

### III. METHODOLOGY TO EVALUATE SENSOR CONFIGURATIONS

We introduce a methodology to evaluate sensor configurations with a constrained Design Space Exploration (DSE) as a tool to support real system design of human motion

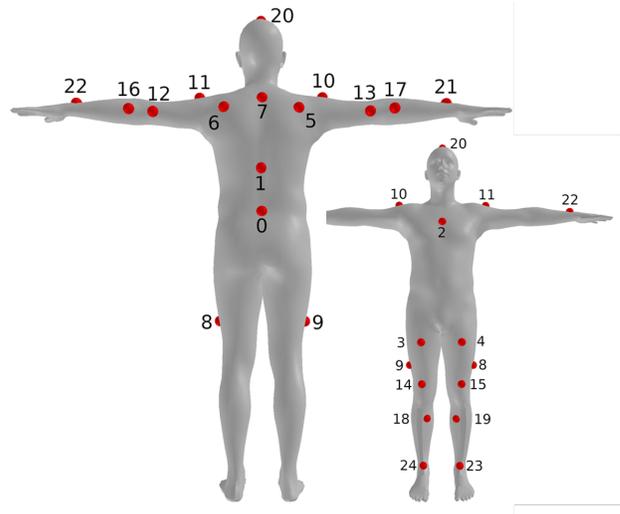


Fig. 3. **Basic sensor configuration** illustrating all possible sensor positions considered for the DSE.

capture from sparse IMU, and propose a combined metric to assess both accuracy and hardware resources. For illustration of this methodology, we exemplarily go through the design of a motion tracking system in rehabilitation to monitor full-body exercises of stroke patients [1]. In this use case, we focus on monitoring the regain of symmetric performance of the exercises after half-sided paralysis and show how a suitable sensor configuration can be determined.

Therefore, we first define a basic sensor configuration as shown in Figure 3, where the sensor positions are inspired by the well-established configurations by Xsens and Steffan et al. [11]. For full-body capture, we add sensors around every limb joint and extend the sensor configuration to be symmetric, ensuring observability of symmetric execution of the rehabilitation exercises. Our definition of the basic sensor configuration includes 25 sensors, i.e., five sensors per leg, four sensors per arm, one at each shoulder, three along the spine, one at the sternum, and one at the head. The exact sensor positions on the SMPL model are depicted in Figure 3.

To ensure reliable full-body recognition and feasibility in real systems, we restrict the maximum number of used sensors to 10 out of the 25 sensors. Otherwise, this would lead to a combinatorial explosion of over 7 million sensor configurations. To limit the considered sensor setups and therefore the computational effort of the DSE, we apply further constraints and reduce redundant caption of the segments by allowing only one sensor per segment, fixing the pelvis node with ID 0 as the root node and focusing on symmetric sensor configurations imposed by the use case. These restrictions lead to a total of 2,249 configurations to evaluate.

For each sensor configuration, we synthesize IMU-data from a body-model dataset using the methods of DIP [5]. For training, we use synthetic data from AMASS, for fine-tuning, synthetic data from DIP-IMU subjects 1 to 8, and for testing synthetic data from DIP-IMU subjects 9 and 10. A holdout

TABLE I

**ERROR COMPARISON ON DIP-IMU DATASET.** FOR EQUAL IMPORTANCE OF ACCURACY AND RESOURCES, WE IDENTIFY A SENSOR SETUP WITH TWO SENSORS LESS THAN THE DIP REFERENCE CONFIGURATION AND IMPROVE SIP BY 10.9%, MESH BY 32.7% AND JITTER BY 82.1%, RESPECTIVELY.

Method	SIP Err (deg)	Mesh Err (cm)	Jitter ( $\frac{\text{km}}{\text{s}^3}$ )
SIP (offline) [4]	21.02 <sup>a</sup>	7.71 <sup>a</sup>	0.38 <sup>a</sup>
DIP [5]	17.10 <sup>b</sup>	8.96 <sup>b</sup>	3.01 <sup>a</sup>
TransPose [6]	16.68 <sup>b</sup>	7.09 <sup>b</sup>	1.46 <sup>b</sup>
PIP [7]	15.02 <sup>b</sup>	5.95 <sup>b</sup>	<b>0.24<sup>b</sup></b>
DIP (ours, 6) <sup>c</sup>	<b>13.63</b>	<b>5.89</b>	0.49
DIP (ours, 4) <sup>d</sup>	15.24	6.03	0.54
DIP (ours, full) <sup>e</sup>	2.77	1.91	0.47

<sup>a</sup>values are taken from [6], <sup>b</sup>values are taken from [7]

<sup>c</sup>DIP reference sensor configuration with 6 sensors

<sup>d</sup>best sensor configuration  $M_4(0.5)$  in terms of mesh error

<sup>e</sup>full basic sensor configuration with 25 sensors

of five sequences from DIP-IMU subjects 1 to 8 is reserved as a validation dataset. For sensor configurations that match the real IMU recordings in DIP-IMU, we proved feasibility of using synthetic data and validated the synthesis approach. We further add white noise during synthesis to map statistical errors of real IMU-sensors and could therefore improve HPE compared to synthesis of ideal sensors.

Since the focus of our work is on the evaluation of different sensor configurations, we take the DNN-model from [5] as is, and train it for each sensor subset. The training procedure consists of 50 epochs of training and 20 epochs of fine-tuning to achieve state-of-the-art accuracy and limit training time.

We log the positional, angular, SIP- and mesh error as well as jitter for each sensor combination. As positional and angular errors are also expressed in the mesh error, we only report the mesh error and therefore still allow for comparison to [4-7]. The SIP-error is the mean orientation error of a joint-subset, i.e., elbow and knee. The jitter evaluates the jerk over a sequence of time frames and therefore estimates the smoothness of the estimation. All metrics are defined as L2-Norms and we average each metric over all sequences of the test dataset.

To evaluate the accuracy-resource trade-off, we propose a combined metric  $M_i$  given in Equation 1, where  $\lambda$  is a weight for hardware resources in relation to accuracy  $e_i$ , e.g. mesh error. For assessment of hardware resources, we approximate costs like energy consumption, bandwidth, latency or wiring of a real sensor setup, by the number of sensors  $i$ .

$$M_i(\lambda) = e_i \cdot (1 - \lambda) + \lambda \cdot i = e_i + \lambda \cdot (i - e_i) \quad (1)$$

$\lambda$  is a design parameter and has to be a value between 0 and 100%, where 0% corresponds to a system with a high emphasis on accuracy and 100% to a high emphasis on hardware resources, respectively.  $e_i$  has to be given in the same order of magnitude as the number of sensors to ensure comparability and avoid an additional scaling factor, i.e., mesh error has to be given in cm without unit. The

TABLE II

**BEST SENSOR CONFIGURATIONS IDENTIFIED FOR MESH ERROR.** THE ERROR RANGE IS GIVEN FOR THE FIVE BEST CONFIGURATIONS. FOR TWO SENSORS, ALL POSSIBLE CONFIGURATIONS ARE LISTED.

Number Sensors	Best configurations	Error range (cm)
2	[2, 0], [7, 0], [1, 0], [20, 0]	12.80 - 13.50
3	[16, 17, 0], [21, 22, 0], [12, 13, 0], [10, 11, 0], [5, 6, 0]	7.55 - 9.11
4	[2, 16, 17, 0], [7, 16, 17, 0], [1, 16, 17, 0], [7, 21, 22, 0], [1, 21, 22, 0]	6.03 - 6.66
5	[5, 6, 21, 22, 0], [5, 6, 16, 17, 0], [2, 20, 16, 17, 0], [1, 20, 16, 17, 0], [2, 20, 21, 22, 0]	5.59 - 5.79
6	[7, 12, 13, 21, 22, 0], [2, 10, 11, 21, 22, 0], [7, 10, 11, 21, 22, 0], [20, 5, 6, 16, 17, 0], [7, 16, 17, 18, 19, 0]	4.92 - 5.01
7	[2, 20, 12, 13, 21, 22, 0], [2, 20, 10, 11, 16, 17, 0], [7, 20, 12, 13, 21, 22, 0], [7, 20, 16, 17, 23, 24, 0], [2, 20, 10, 11, 21, 22, 0]	4.40 - 4.45
8	[2, 10, 11, 18, 19, 21, 22, 0], [2, 10, 11, 16, 17, 23, 24, 0], [2, 12, 13, 21, 22, 23, 24, 0], [20, 5, 6, 18, 19, 21, 22, 0], [2, 10, 11, 14, 15, 16, 17, 0]	3.80 - 3.90
9	[2, 20, 12, 13, 18, 19, 21, 22, 0], [7, 20, 12, 13, 18, 19, 21, 22, 0], [2, 20, 10, 11, 18, 19, 21, 22, 0], [2, 20, 12, 13, 21, 22, 23, 24, 0], [7, 20, 10, 11, 18, 19, 21, 22, 0]	3.22 - 3.28
10	[1, 2, 20, 3, 4, 12, 13, 21, 22, 0], [2, 5, 6, 12, 13, 18, 19, 21, 22, 0], [1, 5, 6, 10, 11, 21, 22, 23, 24, 0], [1, 5, 6, 12, 13, 18, 19, 21, 22, 0], [2, 5, 6, 12, 13, 21, 22, 23, 24, 0]	3.21 - 3.60

optimal configuration corresponds to the configuration with minimal  $M_i(\lambda)$ .

#### IV. RESULTS AND DISCUSSION

For analysis of our methodology, we first describe some general findings and then go into detail of the Design Space Exploration (DSE).

As shown in Table I, due to additional white noise in the synthetic IMU-data, we improve the SIP and mesh error compared to the DIP [5] reference configuration by 9.3% and 34.4%, respectively. The smoothed synthesis with  $n = 4$  frames reduces the jitter by 83.7%. Compared to [5], our full basic configuration of all 25 sensors improves the SIP error by 83.8%, mesh error by 78.4% and jitter by 84.4%, compared to [7], only jitter is degraded.

However, the high number of sensors in the basic configuration does not consider any hardware-awareness. Therefore, we analyze the error metrics results of our DSE depicted in Figure 4. For SIP- and mesh error, many sensor configurations are more accurate than the state-of-the-art implementations from [4-7]. The jitter matches the scale of the two-step methods, which optimize their first prediction with kinematic

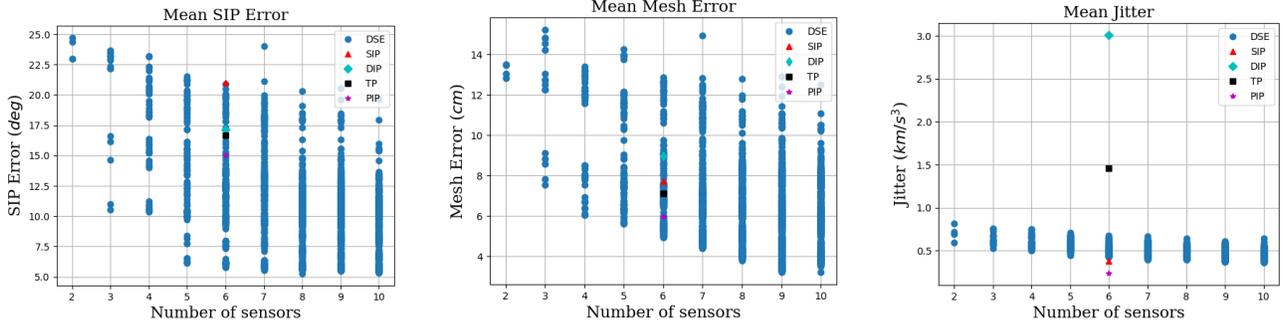


Fig. 4. Accuracy of the sensor configurations for different numbers of sensors.

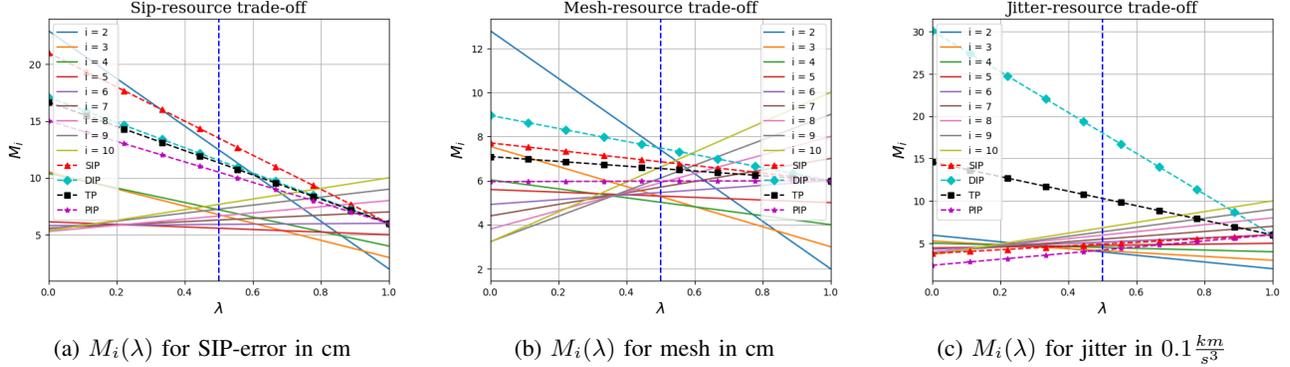


Fig. 5. Combined metric  $M_i$  illustrating the accuracy-resource trade-off on varying hardware weight  $\lambda$  for most accurate sensor configuration of  $i$  sensors. The jitter is scaled by 0.1 to ensure the error metric being at the same order of magnitude as the number of sensors. The vertical blue line indicates a design with equal weight on prediction performance and hardware costs ( $\lambda = 50\%$ ).

constraints like SIP, TransPose and PIP. This originates from the additional white noise in our smoothed synthetic data.

For each error metric, we present the five best sensor configurations in Table II, and count the occurrence of each sensor in the five best and worst setups, see Table III.

The sensor pair with ID 21 and 22, i.e., wrist, occurs the most frequently in the most accurate configurations and rarely appears in inaccurate configurations. Consequently, the wrists are important positions to place sensors. Similarly, the elbow sensors with ID 12 and 13 and the head sensor with ID 20 are included in top-ranked configurations and therefore contribute to an accurate Human Pose Estimation (HPE). In contrast, the back sensors with ID 1 and 7 are the most often part of inaccurate configurations, which is most likely because of their redundancy to the root node and the higher occurrence in setups with a low number of sensors with no limb sensor available. Consequently, for an envisaged low number of sensors, positions at the end of the extremities should be preferred over an alignment at the back.

The foot sensors with ID 23 and 24 contribute to many inaccurate configurations. In summary, sensors mounted on the upper body can be found much more often in accurate configurations and sensors mounted at the lower limbs contribute more often to inaccurate configurations. This implies a higher importance of upper-body sensing than lower-limb sensing for an accurate full-body HPE. However, we are not sure if this implies a general rule or results from imbalanced data, which may be influenced by the dominance of upper-body

motions in the dataset. To finally conclude on the importance of upper or lower body sensing, a detailed analysis of the learning dataset and the focus of the exact application is needed. In our exemplary use case of rehabilitation exercises for stroke patients, the restriction to focus sensing on one body-half would depend on the actual exercises and patients.

For a more precise analysis of the accuracy-resource trade-off, we show our combined metric  $M_i(\lambda)$  from Equation 1 for each best sensor configuration with 2 to 10 sensors separately for SIP-, mesh error and jitter in Figure 5. For a low importance of hardware costs  $\lambda \rightarrow 0\%$ ,  $M_i$  equals exactly the HPE accuracy and for high  $\lambda \rightarrow 100\%$ ,  $M_i$  is defined by the number of sensors, respectively. In between, the beneficial sensor configuration depends on the relative importance of accuracy and resources, which has to be defined by a system designer respecting the specific requirements of the application.

In case of an equal importance of accuracy and number of sensors ( $\lambda = 50\%$ ), all our configurations achieve a better SIP-resource trade-off than state of the art. We identify five sensors as beneficial for SIP-, four sensors for mesh- and two sensors for the jitter-resource trade-off, respectively. As the mesh error captures the human body in total, we focus on this error for a final decision and consider the sensor configuration with sensor IDs 0, 2, 16 and 17 (Figure 6) as optimal to fulfill the accuracy-resource trade-off with equal importance of HPE accuracy and related hardware effort in real applications. This configuration achieves a mesh error

TABLE III

OCURRENCES OF A SENSOR IN THE FIVE BEST/WORST CONFIGURATIONS PER SENSOR ID AND METRIC. THE FIRST AND SECOND VALUE REPRESENT THE NUMBER OF OCCURRENCES IN THE BEST AND WORST CONFIGURATIONS, RESPECTIVELY. THE ROOT SENSOR WITH ID 0 IS PRESENT IN EACH CONFIGURATION. DUE TO SYMMETRY, SOME SENSORS OCCUR IN PAIRS. ACCURATE CONFIGURATIONS ARE HIGHLIGHTED IN BOLD AND INACCURATE CONFIGURATIONS WITH A GRAY BACKGROUND, RESPECTIVELY.

ID	SIP	Mesh	Jitter	Limb location
0	44	44	44	Pelvis
1	<b>14</b> / 15	7 / <b>16</b>	5 / <b>32</b>	Back
2	5 / 9	<b>18</b> / 3	7 / <b>19</b>	Sternum
3, 4	5 / 4	1 / 7	4 / 16	Thigh
5, 6	12 / 2	9 / 1	14 / 5	Shoulder
7	7 / <b>26</b>	10 / <b>20</b>	6 / <b>27</b>	Upper back
8, 9	11 / 10	0 / 13	0 / 9	Thigh
10, 11	<b>15</b> / 1	11 / 1	5 / 1	Upper arm
12, 13	<b>22</b> / 14	12 / 14	<b>21</b> / 4	Upper arm
14, 15	<b>14</b> / 10	1 / 11	9 / 6	Thigh
16, 17	4 / 9	13 / 11	<b>15</b> / 3	Elbow
18, 19	0 / 12	9 / 10	13 / 11	Shin
20	12 / 14	<b>17</b> / 7	<b>16</b> / 10	Head
21, 22	10 / 1	<b>24</b> / 2	7 / 3	Wrist
23, 24	0 / <b>17</b>	6 / <b>19</b>	7 / 10	Ankle

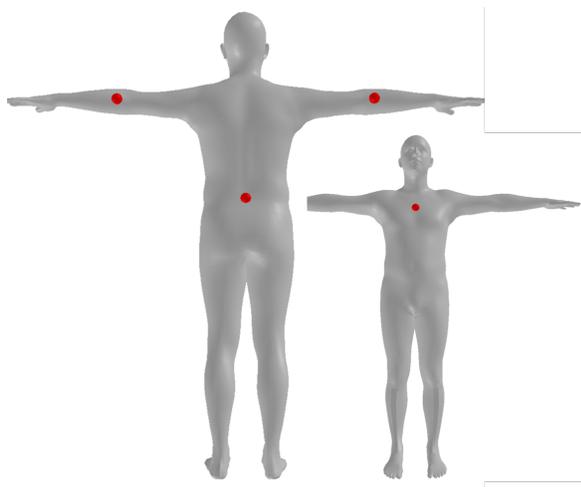


Fig. 6. Resulting optimal sensor configuration for equal importance of accuracy and related hardware effort.

of 6.03 cm reducing the reference of DIP by 32.7% and increasing PIP by 1.3%, while using two sensors less.

## V. CONCLUSION

In summary, we developed a methodology to perform Design Space Exploration (DSE) for Human Pose Estimation (HPE) from sparse IMU-sensing with regard to both prediction performance and hardware costs in real systems and exemplary showed its application in rehabilitation. We therefore synthesized IMU-data from a body model dataset for different sensor setups, trained a deep learning model with this data and evaluated over 2,000 different configurations. Within this DSE, we observed more accurate HPEs with sensors placed on the upper body than on the lower limbs. In

applications with a favored low number of sensors, sensors should rather be placed on the limbs than in line at the back. For assessment of the important accuracy-resource trade-off, we proposed a combined metric with variable importance of both prediction accuracy and hardware resources defined by system requirements and usable by engineers to improve the evaluation of their system. Applying this metric, we identified a sensor network of four sensors at the pelvis, sternum and elbows as beneficial for a system with equal importance of accuracy and resources, resulting in a mesh error of 6.03 cm, which improves state of the art by 32.7% and reduces the hardware effort from six to four sensors. In the future, we will improve our methodology to support the development of real motion tracking systems with fabric-integrated IMU-sensors and attention to data privacy and resource-awareness in diverse health applications.

## REFERENCES

- [1] D. Avola, L. Cinque, G. L. Foresti, and M. R. Marini, "An interactive and low-cost full body rehabilitation framework based on 3d immersive serious games," *Journal of Biomedical Informatics*, vol. 89, 2019.
- [2] F. Kreß, J. Hofer, T. Hotfilter, I. Walter, V. Sidorenko, T. Harbaum, and J. Becker, "Hardware-aware partitioning of convolutional neural network inference for embedded ai applications," in *International Conference on Distributed Computing in Sensor Systems (DCOSS)*, 2022.
- [3] J. E. van Schaik and N. Dominici, "Motion tracking in developmental research: Methods, considerations, and applications," *Progress in Brain Research*, vol. 254, 2020.
- [4] T. Von Marcard, B. Rosenhahn, M. J. Black, and G. Pons-Moll, "Sparse inertial poser: Automatic 3d human pose estimation from sparse imus," in *Computer Graphics Forum*, vol. 36, no. 2. Wiley Online Library, 2017.
- [5] Y. Huang, M. Kaufmann, E. Aksan, M. J. Black, O. Hilliges, and G. Pons-Moll, "Deep inertial poser: Learning to reconstruct human pose from sparse inertial measurements in real time," *ACM Transactions on Graphics (TOG)*, vol. 37, no. 6, 2018.
- [6] X. Yi, Y. Zhou, and F. Xu, "Transpose: Real-time 3d human translation and pose estimation with six inertial sensors," *ACM Transactions on Graphics (TOG)*, vol. 40, no. 4, 2021.
- [7] X. Yi, Y. Zhou, M. Habermann, S. Shimada, V. Golyanik, C. Theobalt, and F. Xu, "Physical inertial poser (pip): Physics-aware real-time human motion tracking from sparse inertial sensors," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.
- [8] N. Mahmood, N. Ghorbani, N. F. Troje, G. Pons-Moll, and M. J. Black, "Amass: Archive of motion capture as surface shapes," in *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019.
- [9] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black, "Smpl: A skinned multi-person linear model," *ACM Transactions on Graphics (TOG)*, vol. 34, no. 6, 2015.
- [10] P. Puchert and T. Ropinski, "Human pose estimation from sparse inertial measurements through recurrent graph convolution," in *Computing Research Repository*, abs/2107.11214, 2021.
- [11] L. Steffan, L. Kaul, and T. Asfour, "Online stability estimation based on inertial sensor data for human and humanoid fall prevention," in *IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, 2017.
- [12] Ö. Terlemez, S. Ulbrich, C. Mandery, M. Do, N. Vahrenkamp, and T. Asfour, "Master motor map (mmm)—framework and toolkit for capturing, representing, and reproducing human motion on humanoid robots," in *IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, 2014.
- [13] C. Mandery, Ö. Terlemez, M. Do, N. Vahrenkamp, and T. Asfour, "The kit whole-body human motion database," in *IEEE International Conference on Advanced Robotics (ICAR)*, 2015.
- [14] C. Mandery, Ö. Terlemez, M. Do, N. Vahrenkamp, and T. Asfour, "Unifying representations and large-scale whole-body motion databases for studying human motion," *IEEE Transactions on Robotics*, vol. 32, no. 4, 2016.

- [15] I. Patzer and T. Asfour, "Minimal sensor setup in lower limb exoskeletons for motion classification based on multi-modal sensor data," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019.