

Figure 11 The twelve subjects of the MHAD dataset.

2.4.1.2 Dataset & evaluation metrics

The comparative evaluation and comparison of the three human pose estimation/tracking methods was based on the Berkeley Multimodal Human Action Database (MHAD) [51]. This dataset features 12 human subjects (see Figure 11). From this figure it can be verified that the MHAD data set involves subjects of considerable variability with respect to age, size and body types. This is also shown quantitatively in Table, columns (G), which provide the lengths of body parts for all the subjects.

The subjects perform 11 different activities (01-jumping, 02-jumping jacks, 03-bending, 04-punching, 05-waving two hands, 06-waving one hand, 07-clapping, 08-throwing, 09-sit down/stand up, 10-sit down and 11-stand up). In each sequence, each activity is repeated several times. The activities are recorded with a multi-camera setup consisting of several conventional cameras as well as by two extrinsically calibrated Kinect sensors. In all experiments reported in this paper, the methods employ the RGBD feeds (both of them for the *HYBRID* method and the same, single feed for the *FHBT* and *OpenNI* methods). The resulting tracking results are compared against the ground truth resulting from the motion capture data.

To quantify the accuracy in body pose estimation, we adopt the metric used in [52]. More specifically, the Euclidean distance between a set of corresponding 3D points (skeleton joints) in the ground truth and in the estimated body model is measured. Each such point (four per leg, three per arm and one for the head) is marked with a red cross in Figure 10. The average of all these distances over all the frames of the sequence constitutes the resulting error estimate Δ .

Another metric reports the percentage $A(t)$ of these distances that are within some predefined threshold t for a certain sequence. We will refer to this metric as the accuracy in human body pose estimation. For example, an accuracy of $A(10) = 70\%$ for a sequence means that in the frames of that sequence, 70% of the joints have been estimated within 10cm from the ground truth.

| Subject | S 01 | | | S 02 | | | S 03 | | | S 04 | | | S 05 | | | S 06 | | |
|---------------------|------|----|----|------|----|----|------|----|----|------|----|----|------|----|----|------|----|----|
| | G | O | F | G | O | F | G | O | F | G | O | F | G | O | F | G | O | F |
| U B L (H I) | 26 | 19 | 22 | 30 | 21 | 24 | 33 | 21 | 23 | 29 | 20 | 22 | 32 | 22 | 23 | 28 | 20 | 24 |
| L B L (I J) | 15 | 19 | 22 | 17 | 21 | 24 | 18 | 21 | 23 | 17 | 20 | 22 | 19 | 22 | 23 | 17 | 20 | 24 |
| S N D (C H, C iH) | 19 | 15 | 17 | 19 | 15 | 17 | 17 | 14 | 17 | 15 | 15 | 17 | 17 | 16 | 17 | 19 | 14 | 17 |
| H N D (G H) | 20 | 25 | 23 | 20 | 25 | 23 | 20 | 25 | 23 | 20 | 21 | 21 | 20 | 26 | 26 | 20 | 20 | 22 |
| L H D (F J, F iJ) | 10 | 9 | 10 | 11 | 9 | 10 | 10 | 8 | 10 | 9 | 9 | 10 | 9 | 9 | 10 | 9 | 8 | 10 |
| B A L (B C, B iC i) | 24 | 25 | 28 | 28 | 27 | 31 | 31 | 28 | 33 | 24 | 23 | 25 | 26 | 28 | 33 | 26 | 26 | 30 |
| F A L (A B, A iB i) | 23 | 26 | 21 | 25 | 31 | 29 | 26 | 32 | 29 | 24 | 25 | 23 | 25 | 31 | 28 | 24 | 27 | 25 |
| B L L (E F, E iF i) | 36 | 41 | 39 | 43 | 47 | 41 | 44 | 47 | 41 | 37 | 39 | 39 | 42 | 45 | 40 | 42 | 44 | 42 |
| F L L (D E, D iE i) | 42 | 37 | 39 | 48 | 42 | 41 | 47 | 43 | 41 | 41 | 35 | 40 | 45 | 42 | 40 | 45 | 41 | 41 |

(a)

| Subject | S 07 | | | S 08 | | | S 09 | | | S 10 | | | S 11 | | | S 12 | | |
|---------------------|------|----|----|------|----|----|------|----|----|------|----|----|------|----|----|------|----|----|
| | G | O | F | G | O | F | G | O | F | G | O | F | G | O | F | G | O | F |
| U B L (H I) | 25 | 20 | 21 | 30 | 20 | 24 | 27 | 20 | 21 | 27 | 21 | 23 | 28 | 20 | 23 | 24 | 20 | 23 |
| L B L (I J) | 15 | 20 | 21 | 18 | 20 | 23 | 15 | 20 | 21 | 16 | 21 | 23 | 16 | 20 | 23 | 21 | 20 | 23 |
| S N D (C H, C iH) | 17 | 17 | 17 | 18 | 15 | 16 | 15 | 13 | 17 | 17 | 14 | 17 | 17 | 15 | 17 | 18 | 14 | 17 |
| H N D (G H) | 20 | 20 | 21 | 20 | 21 | 20 | 20 | 17 | 23 | 20 | 24 | 21 | 20 | 25 | 22 | 20 | 24 | 22 |
| L H D (F J, F iJ) | 8 | 7 | 10 | 9 | 9 | 10 | 8 | 7 | 10 | 8 | 8 | 10 | 8 | 9 | 10 | 9 | 8 | 10 |
| B A L (B C, B iC i) | 22 | 25 | 26 | 24 | 26 | 28 | 23 | 22 | 24 | 27 | 25 | 30 | 26 | 27 | 31 | 25 | 25 | 27 |
| F A L (A B, A iB i) | 22 | 24 | 23 | 24 | 26 | 24 | 23 | 27 | 22 | 24 | 29 | 24 | 24 | 28 | 24 | 22 | 27 | 24 |
| B L L (E F, E iF i) | 35 | 39 | 38 | 39 | 40 | 41 | 35 | 42 | 38 | 41 | 43 | 40 | 41 | 44 | 41 | 38 | 43 | 41 |
| F L L (D E, D iE i) | 41 | 35 | 37 | 43 | 39 | 41 | 41 | 37 | 40 | 43 | 40 | 40 | 44 | 44 | 41 | 41 | 39 | 41 |

(b)

Table 9: Body part lengths (in cm) for the human subjects of the MHAD dataset, (a) subjects 01-06, (b) subjects 07-12. Columns (G) are the manually measured, ground truth values, columns (O) the one estimated by the *OpenNI* method, and columns (F) are the ones estimated by the *FHBT* method. The parenthesis next to the name of each body part to the corresponding body segment(s) in Figure 9.

2.4.1.3 Comparative evaluation

Several experiments were carried out to assess quantitatively the accuracy and the performance of the evaluated human articulation tracking methods.

All-subjects-one-action experiment: A first experiment aimed at evaluating the performance of the methods across different human subjects. All twelve sequences showing the twelve different subjects performing the same activity (activity 04, boxing) were considered. It has to be noted that the *FHBT* method estimates only a subset of the joints, depending on whether the method's confidence on them exceeds an internally set threshold. Figure 12 (c) shows the percentage of joints that were estimated by each method. For the *OpenNI* and *HYBRID* methods this is always 100% while for the *FHBT* method this is 75% on average, across different subjects. In a subsequent measurement, we evaluated the error Δ and the accuracy $A(10\text{cm})$ (Figure 12 (d), (e), respectively) for all methods, but only over the joints and the frames for which the *FHBT* method provided some estimation. It can be verified that when error and accuracy is measured over these joints, the performance of the *OpenNI* and the *HYBRID* method increases. This indicates that the confidence that *FHBT* estimates for the joints is trustworthy.

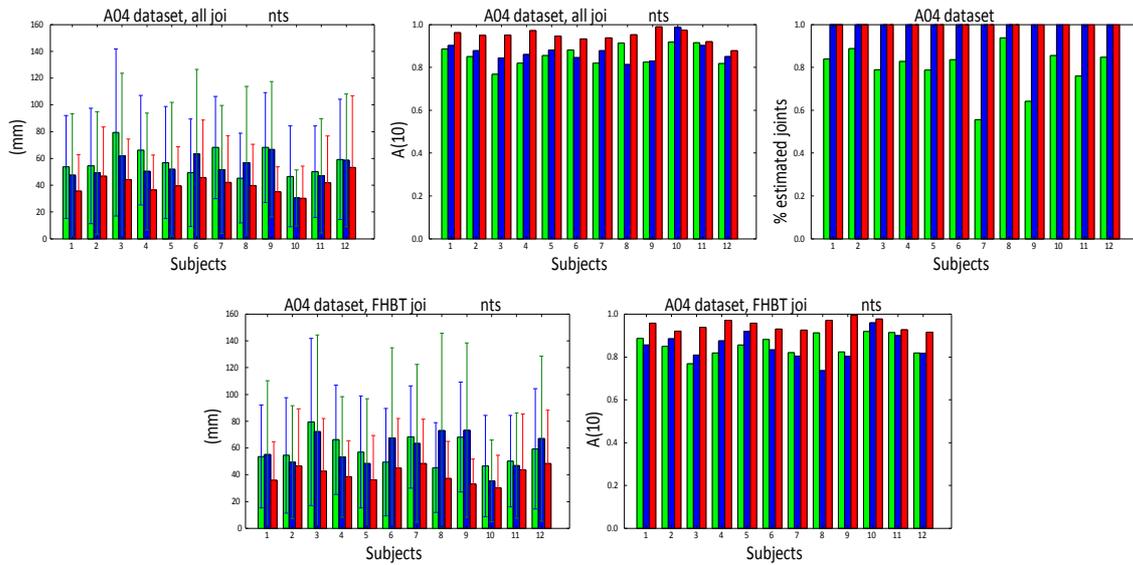


Figure 12: Quantitative evaluation of the method applied to 12 subjects performing the same action (boxing). (top-left) Error Δ and variances over all frames and joints. (top-middle) Accuracy $A(10\text{cm})$ over all frames and joints. (top-right) Percentage of joints for which a method provided an estimation. (bottom-left), (bottom-right): Error Δ and accuracy $A(10\text{cm})$ over the joints for which *FHBT* provided an estimation. *FHBT*: green bars, *HYBRID*: red bars, *OpenNI*: blue bars.

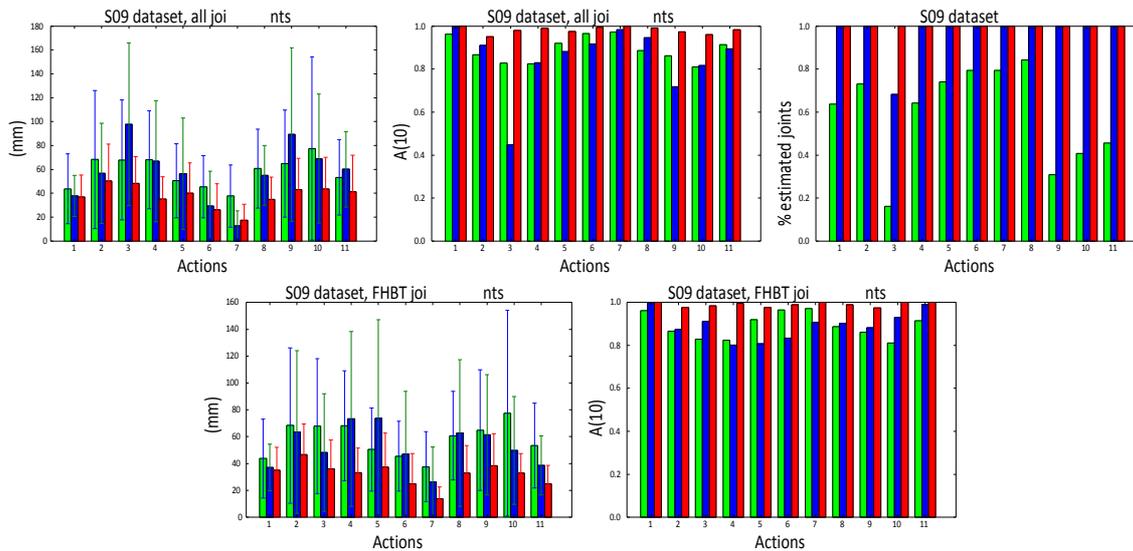


Figure 13: Quantitative evaluation of the method applied to 11 actions performed by the same subject (s09). (top-left) Error Δ and variances over all frames and joints. (top-middle) Accuracy $A(10\text{cm})$ over all frames and joints. (top-right) Percentage of joints for which a method provided an estimation. (bottom-left), (bottom-right): Error Δ and accuracy $A(10\text{cm})$ over the joints for which *FHBT* provided an estimation. *FHBT*: green bars, *HYBRID*: red bars, *OpenNI*: blue bars.

All-actions-one-subject experiment: In a second experiment, the goal was to assess the proposed method with respect to different activities. For that purpose, the evaluation was performed on image sequences showing a single subject performing the eleven different activities. Figure 13 illustrates the obtained results in a way analogous to that of Figure 12. Again, *HYBRID* outperforms the rest of the methods with respect to the mean error Δ and accuracy, while the rest two methods perform comparably. It should also be noted that for actions like bending (action 03) and sit-down/stand-up (action 09) that exhibit considerable self- and body-object occlusions, the *FHBT* method estimates the least number of joints (see Figure 13 (c)).

Aggregated results: Table 10 summarizes the performed experiments by providing Δ , $A(10)$ numerical values for the cases of all-subjects-one-action and all-actions-one-subject experiments, as well as for the union of the corresponding datasets. A number of interesting conclusions can be drawn: (a) Overall, the *HYBRID* method is the one that results in the lowest errors and error variances and the highest accuracy, (b) *FHBT* and *OpenNI* perform comparably, (c) *FHBT* exhibits minimum performance variability between the two experiments with respect to Δ and its standard deviation, while *OpenNI* exhibits minimum performance variability with respect to $A(10)$ and, (d) *HYBRID* has the maximum variability in all metrics. In the results of Figure 12 and Figure 13 the accuracy $A(t)$ has been computed for $t=10$ cm. While this choice of t is compatible with the requirements of many applications, it is interesting to know how the accuracy of a certain method varies as a function of t . Figure 14 presents this information. For the three evaluated methods, we measure their accuracy for various values t in the range $[0..20]$ cm. The top row of plots shows these results over the joints that the *FHBT* method estimated, while the bottom row shows the same results over all the joints. It can be verified that the *HYBRID* method is consistently more accurate compared to the other two, regardless of t . *FHBT* and *OpenNI* perform comparably. Moreover, the plots show for which error tolerances each method becomes preferable.

Estimation of body sizes: The *HYBRID* method relies on its discriminative part (which is the *OpenNI* method) in order to initialize the tracking process and to set the proper human body model parameters. The *FHBT* method has its own mechanism to provide an estimation of these parameters. Table 9: shows, for each subject, the ground truth information (columns (G)) as well as the ones estimated by the *OpenNI* (columns (O)) and *FHBT* (columns (F)) methods. It can be verified that the *FHBT* method is slightly more accurate in estimating body shape parameters than *OpenNI*. In particular, for each method, we computed the mean error (among subjects) in the estimation of each body part length. Then, we estimated the mean and the standard deviation of these errors for all body parts. The results show that the mean error in the estimation of the body parts for the *OpenNI* method is 3.44cm with a standard deviation of 0.68cm, while for the *FHBT* method we obtain a mean error of 2.86cm with a standard deviation of 0.71cm. The analysis also shows that the most inaccurate measurements are obtained for the human torso-related parts, while the lengths of the limb parts (arms, legs) are estimated more accurately.

| Dataset | S 09 | | | A 04 | | | A ggregate | | |
|------------|-----------|-----------|------------|-----------|-----------|------------|------------|-----------|------------|
| | Mean | S td. | A (10) (%) | Mean | S td. | A (10) (%) | Mean | S td. | A (10) (%) |
| F H B T | 58.0/58.0 | 40.7/40.7 | 89.2/89.2 | 58.1/58.1 | 41.3/41.3 | 85.5/85.5 | 57.6/57.6 | 41.0/41.0 | 87.5/87.5 |
| O penNI | 69.4/52.8 | 63.1/50.1 | 80.7/89.4 | 67.6/58.7 | 69.5/58.9 | 80.1/84.9 | 67.9/55.1 | 66.3/54.2 | 80.6/87.4 |
| H Y B R ID | 36.1/32.3 | 20.3/19.0 | 98.5/98.9 | 42.6/40.5 | 34.7/32.7 | 93.5/94.9 | 39.7/36.7 | 28.3/26.5 | 95.8/96.7 |

Table 10: Comparison of *FHBT*, *HYBRID* and *OpenNI* methods in all datasets. Mean Δ and std. of Δ are measured in mm. The two numbers in each slot of the matrix refer to the quantity measured over all joints/the quantity measured over the joints computed by *FHBT*.

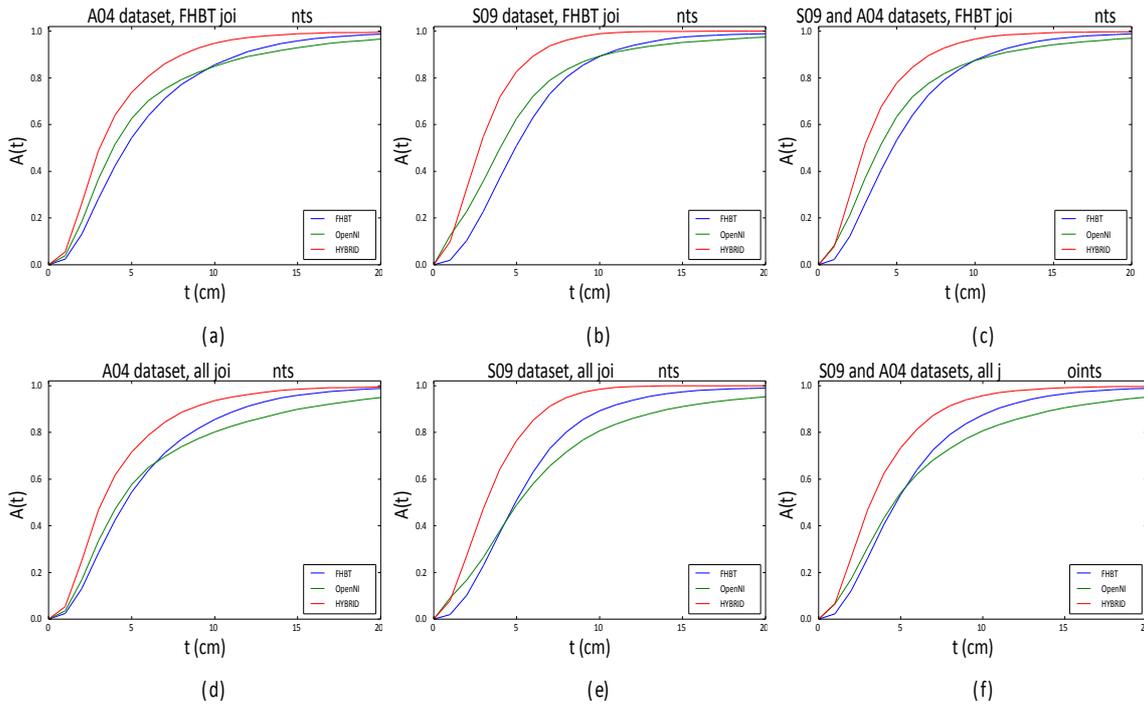


Figure 14: The accuracy $A(t)$ of the evaluated methods as a function of t in the range $[0..20]$ cm, for all experiments. Left column: all-subjects-one-action, middle column: all-actions-one-subject, right column: the union of the two datasets. Top row: results for the joints estimated by *FHBT*. Bottom row: results for the joints that each individual method estimates.

2.4.1.4 *FHBT* Limitations

A series of experiments performed on a ground-truth-annotated data set demonstrated quantitatively and qualitatively the performance of the evaluated methods. The results show that in situations where small error and high accuracy is more important than the burden and the overhead of using a second RGBD sensor, the *HYBRID* method is the preferred one. Interestingly, the *HYBRID* method is slightly less accurate than other purely generative methods like pPSO [46] that are aware of an accurate human body model. Still, the fact that *HYBRID* is fully automatic, is a significant advantage that, depending on application, might be more important than its lacking accuracy.

Another result is that *FHBT* and *OpenNI* perform comparably. *FHBT* has some additional practical advantages that make it an attractive alternative for estimating human 3D pose. For example, it initializes instantly (in a single frame), can cope with partially visible human bodies and operates with a moving camera, even in jerky motion. It should be stressed that the employed MHAD dataset does not showcase such difficult situations which are, nevertheless, abundant in several real-life scenarios⁷ including the ones considered in the context of the ACANTO project. While the *FHBT* handles self-occlusions and camera motion well, the method has its limitations when trying to track a person that is very close to the camera and thus not fully visible or when there are severe occlusions with the person partially hidden behind objects (i.e. furniture). The main area of operation for the ACANTO *FriWalk* is the indoors environment of the hospital as well as the care facility for the elderly. The walker must be able to function in the limited space of the patient's room and around the typical furniture and obstacles found in the aforementioned type of environments. More specifically the, *FriWalk* assisted, exercises and tests (described in detail in chapter 3) will be executed in patient rooms with very strict space limitations. These tests and exercises are problematic cases for the baseline body tracker which was designed to work in more open spaces.

⁷ See also <https://youtu.be/ZKlC9PA11Dg>

2.4.1.5 Clinical datasets

In order to test the *FriWalk* body tracking modules in realistic settings, a hospital environment is necessary. FORTH has a hospital room simulation and test area in the AMI (Ambient Intelligence) facility in Heraklion, Crete. With the support and information provided by the physicians in Getafe, the hospital room simulation was restricted to the expected dimensions of the patients' rooms in Getafe. Similar lighting conditions as well as the furniture (i.e. hospital bed, chairs) were used. The datasets were acquired with five different subjects, two female and three male. Two different exercises were performed by each test subject: (a) the "Chair stand test" with the person using a chair in the simulation room and (b) the "Isometric/Isotonic exercise" with the person sitting on the clinical-bed in the simulation room. These tests are presented in detail in sections 3.1.3 and 3.1.4 respectively. For each exercise each subject was asked to execute it multiple times in "successful" and "failed" scenarios. The datasets acquired in the simulation facility were used to test and tune the body tracking modules (Figure 15, Figure 16 and Figure 17).

2.4.1.6 Task-specific body tracking

The environment and application constraints gave rise to the need for specialized algorithms and adaptations for the body tracker modules of the *FriWalk*. First, the restrictions in the walker placement (and, as a result, of the camera placement) create complete occlusions in the lower limbs. This is an important issue in the ACANTO setup where accurate position of the legs is required during the clinical tests and exercises. Secondly, the clutter in the environment (bed, sheets, furniture) introduces uncertainties that eventually lead the original *FHBT* module to false or inaccurate detections. Finally, the joint-tracking accuracy required in the ACANTO clinical tests brings the adaptive (but still generic) body model used by the *FHBT* implementation to its limits. FORTH is currently in the final development and testing stages of a number of alternative methodologies that work around and when possible leverage the environmental restrictions to enable robust body tracking.

User-specific model tracking: The approach is using a model that is tailored to the specific user. Initially the user is scanned using an RGBD sensor. The captured point clouds are fused into a single mesh using [47]. The 3D mesh is then being rigged with a properly adjusted skeleton using [48]. The scanned model is used with a model based optimization pipeline which is building upon the success of FORTH's hand tracking framework (Figure 15).

Advantages of this model based technique is (a) the ability to focus on optimizing the accuracy for specific joints and (b) a higher accuracy since the model matches the observed human. Results of this optimization pipeline is shown on Figure 16 alongside results by the *FHBT* for the same frames.

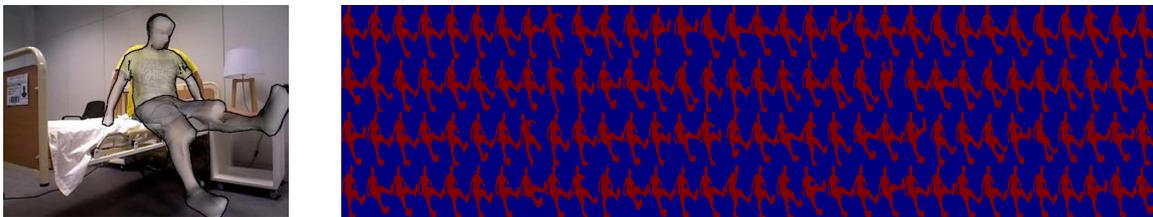


Figure 15: Left: Sample result using a user-specific model, tracking is performed with an adaptation of FORTH's model based tracking framework. Right: evaluated hypotheses (body configurations) generated for the frame on the left using particle swarm optimization, the correct hypothesis is on the top left.

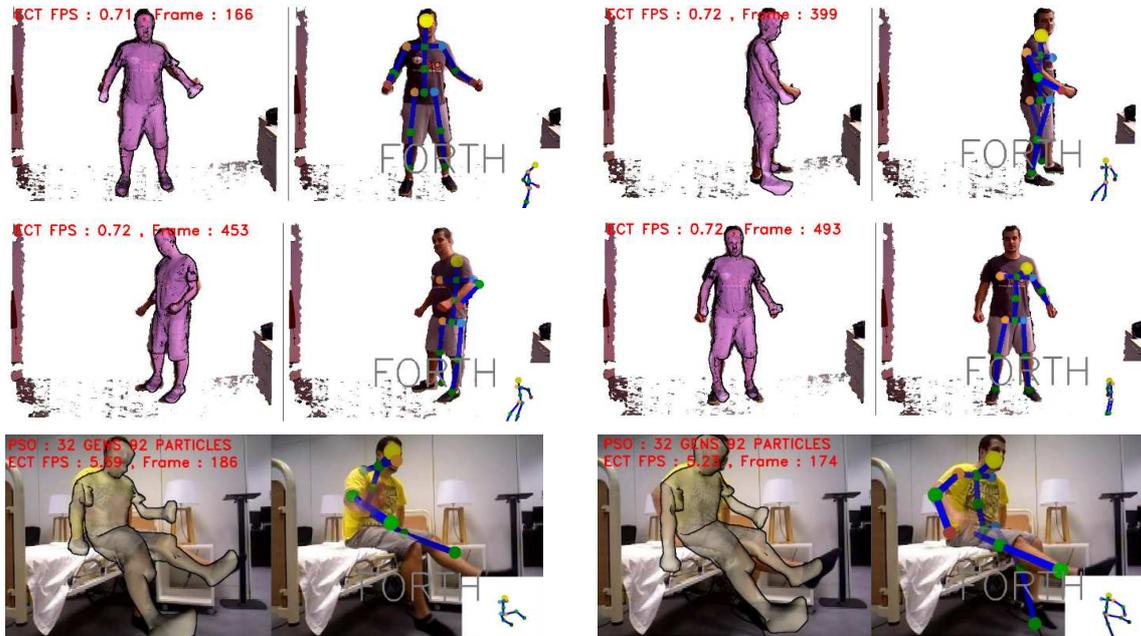


Figure 16: Preliminary tracking results using the body tracker with the personalized model. A walking sequence and a sequence with a person performing an isometric exercise are shown. The first and third columns show the user-specific model results while the second and fourth columns show the FHBT result on the same frame. For the FHBT frames, the side-view of the tracked skeleton is shown super-imposed on the bottom right.

The developed generative method can be used in conjunction with a deep learning approach such as [49] which can provide strong initialization and re-initialization priors. The user-specific model based approach can provide improved accuracy and enable continuous patient tracking even when the lower limbs are completely occluded for a number of frames.

HMF Body Tracking: The previously mentioned human pose tracking based on a personalized human model capitalizes on Particle Swarm Optimization for fitting the human model into the available observations. This optimization framework shows promising results, however has two important drawbacks (a) it maintains and propagates a single hypothesis regarding the configuration of the human body, thus, if tracking fails, it cannot be recovered (b) while there is considerable parallelism at the particles level, all particles generations need to be executed serially. To mitigate these issues an alternative optimization technique based on the hierarchical particle filter was implemented in the “HMF Body Tracking” module.

This approach considers a generative, particle filter (PF) applied to the body tracking problem. Similar to the “User-specific tracking” method, the method relies on a 3D model of the human body. The body has several degrees of freedom that encode its 3D position and orientation as well as the joint angles that are relevant to each tracking scenario. Successful tracking boils down to estimating the state of the model at each frame given the RGB-D observations. A hierarchical particle filter (HMF) [1, 3] that can handle high dimensional state spaces is employed for the task. To ensure robust tracking for persons with different somatometric measurements a model shape adaptation procedure [28] is performed in the beginning of each session. The adaptation procedure adapts the shape of the generic 3d model that we use for initialization to the actual shape of the person that is being tracked.

The HMF tracker follows a hierarchical strategy to estimate the target’s state. A set of auxiliary models that lie in lower dimensional spaces and track parts of the body are used. At each frame these models are updated first and the information from them acts as prior for the main model that is the full body pose. This way the search in the high dimensional space of the main model is narrowed significantly making the method faster and suitable for real time applications. As a

PF variant, HMF uses a set of hypotheses to approximate the posterior state distribution given the observations. To measure the likelihood of each hypothesis we use a combined objective function that computes the weighted average of the following two terms: (i) the pixel per pixel discrepancy between a rendered pose and RGB-D observation. (ii) The distance between the model joint positions and the joint positions as detected by the bottom-up *FHBT* detector. Using these two terms for the calculation of the likelihood results in a tracker with increased robustness that outperforms trackers relying on one modality.

We tested the described tracker on standard datasets as well as on our dataset that contains sequences that correspond to the ACANTO exercise scenarios (see Figure 17). The preliminary qualitative results show that the approach performs well in terms of tracking accuracy and speed (10-30fps depending on the number of particles).



Figure 17: Preliminary results of HMF body tracking of a subject performing the leg extension isometric exercise.

2.5 Mechanical Sensors for User Detection

Since the *FriWalk* is capable of autonomous motion, one of the main risks that is potentially harmful for the user is related to a severe software failure, which could make the robot move unpredictably. The *FriWalk* could in this case leave the user on its own alone, running over the user or causing her/him to lose her/his balance. In order to minimize/eliminate the identified risks and preserve the user safety, we installed two mechanical devices:

- A safety lanyard which connects the user's chest with the robot. If the distance between the user and the robot exceeds the lanyard length, then the power supply is immediately and automatically detached;
- A safety switch mounted in the two bars of the seat. In this way, if the *FriWalk* goes backwards, as soon as the seat comes into contact with the user legs the power supply is immediately and automatically disconnected.

We point out that both mechanisms operate directly on the power switches; therefore they do not need any software signal to be activated. Nonetheless, the use of correct-by-construction programming mechanisms, of well tested versions of the Linux Kernel and of well known computing devices (adopted and supported by a large community) mitigates the risk of a severe software failure.

To address a similar problem, i.e. the user and the *FriWalk* should remain in physical contact during standard navigation tasks, the rollator grips are equipped with a pair of resistive contact sensors, embedded inside the handles. In such a way not only the walker is prevented from moving on its own unless required by the specific application, e.g. for some of the clinical tests foreseen in the next Chapter, but it can also fulfil specific functional purposes. For instance, some clinical exercises require the user to hold the grips to show that her/his attention is focused on the exercise. In addition, in some other exercises the user is explicitly not allowed to use the handles, otherwise the exercise/physical performance test will be rated as a fail.

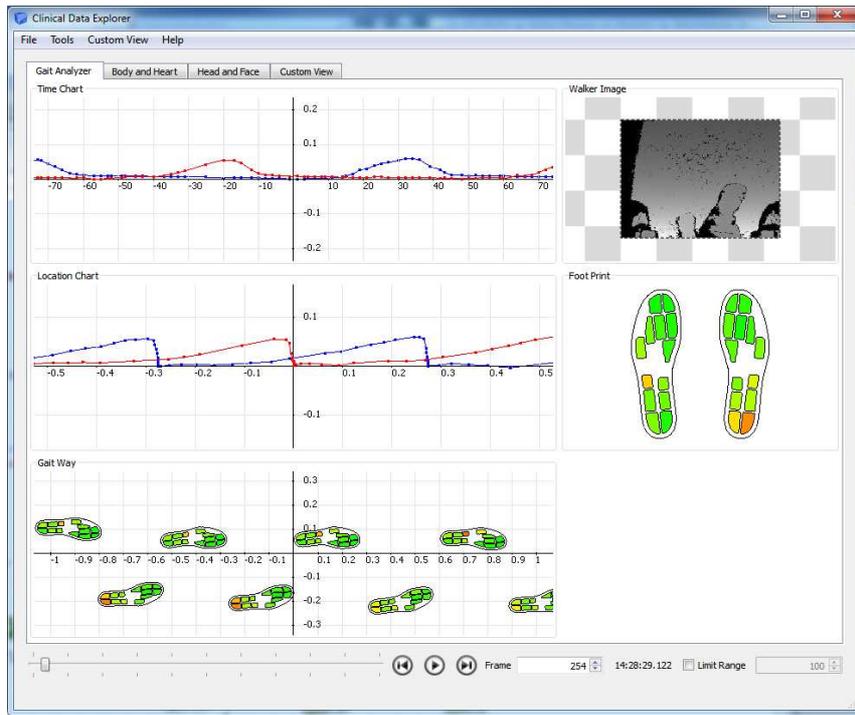
The adopted resistive sensor, precisely the *Force Sensitive Resistor* $L = 600\text{ mm}$ developed by FSR, generates analogic outputs that can be acquired by the embedded computing platform on

board the *FriWalk*. To allocate the sensor, the original rubber handles have been cut and then fixed back.

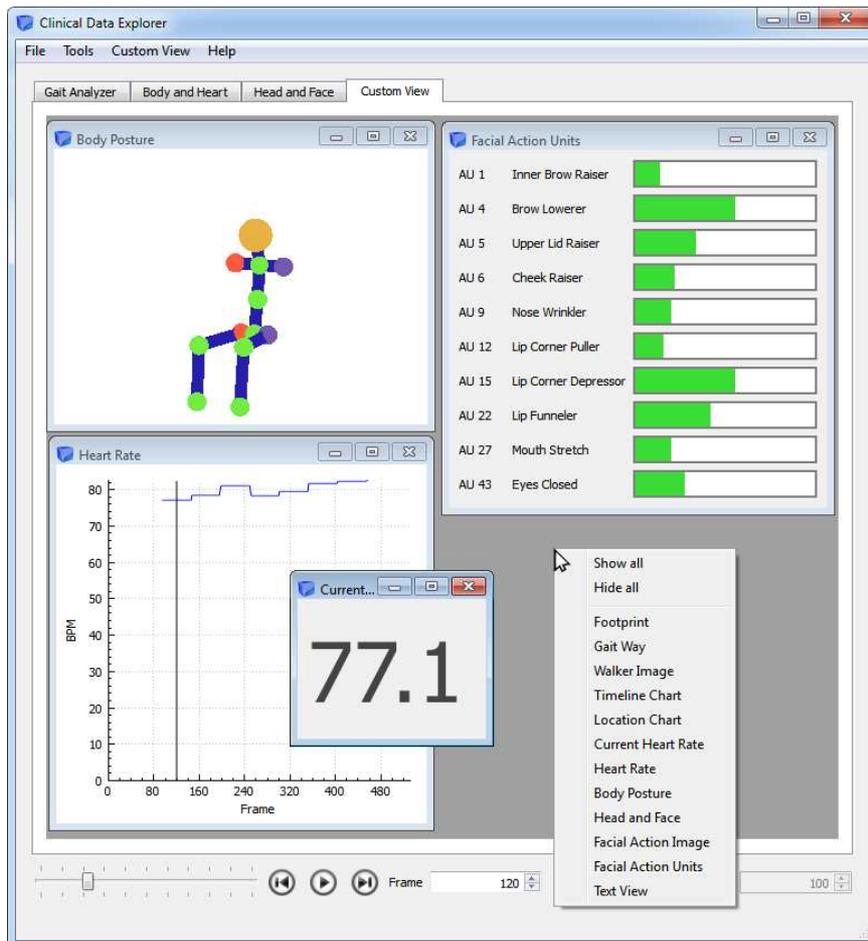
2.6 Clinical Data Explorer

For investigation purposes a graphical user interface (“Clinical Data Explorer”) has been developed that allows for interactive visualization of all the measured user specific quantities. At the moment only the acquired data is shown whereas in the future this tool could also be extended to calculated derived measures like e.g. velocity and acceleration parameters for gait in form of various 2D and 3D plots.

The tool currently consists of predefined views for user state related modalities (Figure 18 (a) shows the view on gait aiming to provide similar information like a gait walkway) as well allows for a so called “custom view” that allows to arrange the user data freely across individual modalities (Figure 18 (b)).



(a)



(b)

Figure 18: Clinical Data Explorer.
 (a) Gait Analyzer View. (b) Custom View.

Chapter 3

Activity Analysis in the Clinical Scenario

For the clinical environment in a workshop held at the University Hospital of Getafe the project consortium identified different categories of activities that benefit from automated activity analysis:

- Diagnostic activities (SPPB test)
- Orthogeriatric exercises
- Training exercises

The rationale behind automation is a higher degree of quantification and hence repeatability as opposed to currently manual measurements taken, as well as the opportunity to execute these tests without the presence of highly qualified medical professionals up to even unsupervised execution of certain exercises.

For that reason keeping the process simple and safe for the older adult is of utmost importance, so the project team did a careful consideration of safety mechanisms e.g. a chord with a magnetic switch is attached to the patient dress as described in 2.5. Should the robot start moving forward, the switch is stripped off and the robot stops. Hence the project team did consider trade-offs between technical feasibility of features of our robotic device (like autonomous movement) and potential risks introduced by it.

In addition since *FriWalk* is a robotic device it can be used also for people that do not need the support of a walker, in the sense that the *FriWalk* is just acting as a mobile measurement device, that the older adult has to follow or that follows the older adult autonomously. This also includes automatic placement (on the longer run) of the walker for certain stationary tests or at least (visual) guidance for aiding a caregiver in correctly placing the walker.

3.1 SPPB Test

The short physical performance battery (SPPB) is a group of measures that combines the results of the gait speed, chair stand and balance tests [1]. It has been used as a predictive tool for possible disability and can aid in the monitoring of function in older people. The scores range from 0 (worst performance) to 12 (best performance). The SPPB has been shown to have predictive validity showing a gradient of risk for mortality, nursing home admission, and disability.

It consists of a standardized sequence of three individual tests as can be seen from Figure 19.

The first test (“Balance Tests”) rates if the older adult is able to keep certain stances for a given period of time without losing the balance. In our case of using a smart walker, in order to successfully complete the test the person is not allowed to use the handles to keep balance, however the presence of the handles offers additional safety just in case. Technically speaking we need to rate the presence of **specific stances derived from our gait analysis system**.

The second test (“Gait Speed Test”) measures the time required to walk 4 meters at a normal place, and it is recommended to use the best time out of two trials. While this test sounds rather simple, user-centric sensing allows us to gather **additional information during the short walk** and yields quantitative measurements not been available up to now with the common practice of taking measurements with a stopwatch.

The last test (“Chair Stand Test”) measures the time required to perform five rises from a chair to an upright position as fast as possible without the use of the arms. With respect to user-

centric sensing the test is based on **automatic human body model analysis** and requires a clever placement of the robotic walker with respect to the older adult performing the test, in order to ease the detection process.

Further details on the individual tests will be given in the subsequent chapters.

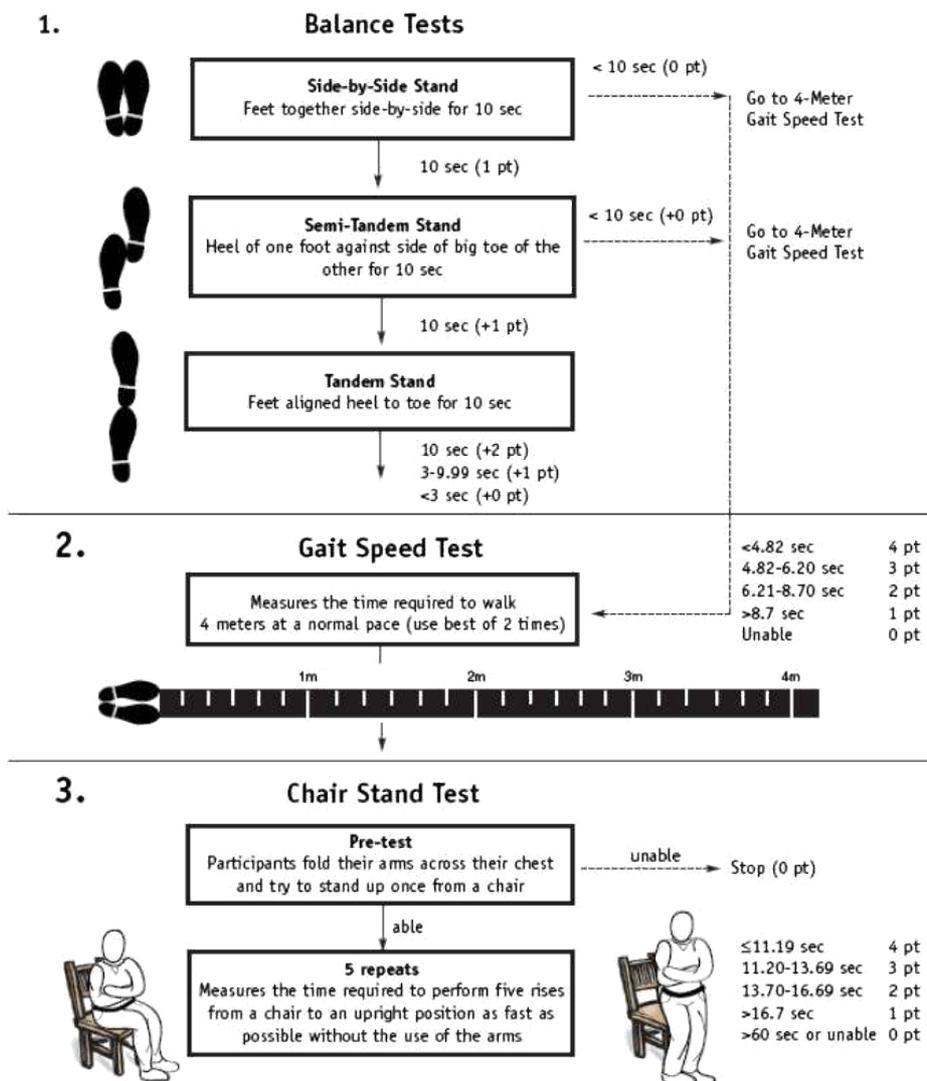


Figure 19: SPPB Test - Standardized test procedure and associated rating system.

3.1.1 Balance Test

The sequence of “Balance Tests” rates if the older adult is able to keep certain stances for a given period of time without losing the balance. Utilizing the messages from the intermediate level representation test sequences like the balance test can very easily be scripted and this lightweight application logic eases the subsequent integration into a graphical user interface. Figure 20 shows a prototype of the balance test for testing and demonstration purposes. The GUI is not yet fully optimized for potential end users. Harmonization in visualization across all individual activity analysis modules in the clinical scenario will take place later.

Please note that the tandem stand can usually only be detected by the circumstance that one foot is visible, so foot colouring is missing for left and right foot respectively.

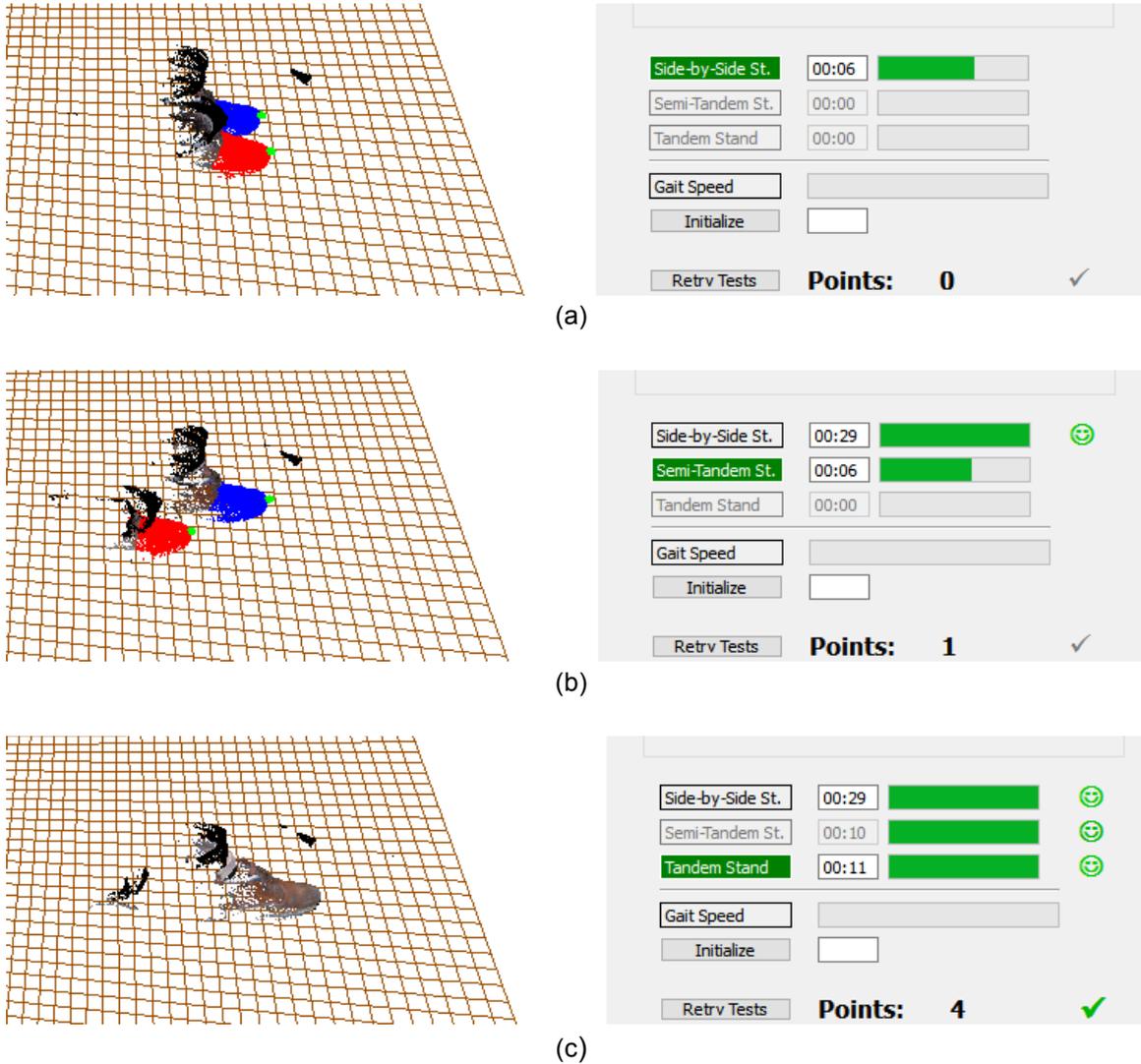


Figure 20: Progression of SPPB test.

Timer bars shows the progress of each test and the performance is automatically rated according to the point system given in Figure 19. (a) Side-by-Side Test. (b) Semi-Tandem Stand. (c) Tandem Stand.

3.1.2 Gait Speed

The user gait speed can be computed directly from the reconstructed *FriWalk* position coming from the localization algorithm, described in Chapter 5. As reported there, the localization algorithm is based on a coherent fusion between dead-reckoning data (i.e. encoders and, in case, visual odometry or IMU data), absolute position data (i.e. from QR landmarks or, in case, cloud positioning systems, such as surveillance cameras) and relative position measures (i.e. collaborative localization). In this setting, to recover a stable and accurate velocity measure two problems should be tackled. First, the user location tracking is noisy; hence the data should be properly filtered. Second, when the position is updated by an absolute reference, the estimated location jumps due to the intermittent availability of such information (more on this in Chapter 5). Even though the collaborative localization approach mitigates this detrimental effect, the problem remains (see Figure 27 in Chapter 5).

To deal with both the aforementioned negative aspects, the gait speed can be computed with a low pass windowed mean filter applied on the estimated user positions, in which the window

length is 250 ms, corresponding to about 60 consecutive position estimates. The length of the filter has been computed to filter out at most both the noise and the jumps related to the absolute updates. Notice that 250 ms are a sufficiently small amount of time to assume that the velocity of the user is almost constant, unless it stops on the spot. To detect such a sudden stop situation, which will generate a slow fading of the measured velocity of length 250 ms due to the filter window, we make use of the encoders, in a way that resembles the Zero velocity Update (ZUP) used for location tracking with foot mounted accelerometers [80]: when the encoders stop, we reset the gait speed to zero. As a final remark, we have to point out that the use of a smoother to further reduce the effect of the jumps in the estimates can be adopted.

Nonetheless, since the gait velocity can be considered as an instantaneous measure, very much similar to the tachometer of a standard ground vehicle, this quantity can be measured using only encoder data captured on the 250 ms aforementioned window and then filtered. This approach is simpler than the localization-based solution and relies on the fact, substantiated by experimental evidence on the field, that the encoders noise is approximately white with zero mean. This approach, currently adopted on the *FriWalk*, still partially makes use of the localization algorithm by removing from the encoder data the drift terms estimated in the filter (see Chapter 5). It is worthwhile that the velocity estimates can be further refined by the angular velocity directly available from the rear motors driver, even if it is not yet implemented at the moment of writing.

3.1.3 Chair stand Test

The chair stand test is part of the SPPB tests in Figure 19. The goal of this test is to measure the frailty of the patient and give a measurement of the probability this patient has to suffer a fall with a possible bone fracture in the near future. During the test the patient is asked to sit on a chair and stand up again while having his or her arms crossed on the chest. The results of the test are determined by the time it takes for the patient to perform each repetition.

The implementation of the test using the *FriWalk*, apart from automating the process, enables the recording of the full body posture of the patient during the test. This information along with the test scores is made available to the physician for further analysis.

The protocol for the execution of the test was created such that it minimizes the risk for the patient and allows non-medical personnel to provide assistance and function as caregivers. The test is split into phases: instructions phase, positioning, pre-test phase and test phase.

On initiation the test starts with the “instructions phase”. The patient puts her/his hands on the *FriWalk* grips and receives audible and visual instructions. The instructions explain the test requirements. Once the patient is ready the test enters the second phase. Then the patient is asked to sit on a chair. In order to provide the person with support as he/she sits the walker is firmly locked (brakes on).

Subsequently the walker is positioned in front of the person in a way that the full body of the patient can be observed by the front-mounted RGB-D sensor. Once the walker is properly positioned the pre-test phase starts. In this step the goal is to test if the patient is capable to perform the actual test procedure. The system gives the audible command “STAND UP and SIT DOWN PLEASE” to make sure that the patient is able to reach a full knee extension with arms crossed. If the patient qualifies she/he is informed that the test is ready to start.

The system says “SIT DOWN” and then “GO”. If, during the test, the patient moves his/her arms away from the correct pose (crossed on the chest) or loses balance, he/she fails and the test stops. If the patient fails to complete the requested number of repetitions in 1 minute the test fails. Each time a successful sit-stand repetition is completed the system emits an audible “tick” giving feedback to the patient and the assistant. The system records the time that the patient

spends carrying out the 5 repetitions (using the scoring system), or if the patient is unable to do the exercise. The walker autonomously detects the end of the test. Finally the walker is returned in the original position to help the patient to stand-up.

3.1.4 Isometric/Isotonic Exercise

The goal of the isometric/isotonic exercise is to strengthen the leg muscles of the patient. This type of exercise is typically prescribed by physicians for patients that live in nursing homes or are recovering in a clinic. The patient is expected to perform a number of knee-bending repetitions while seated. Each repetition is timed. The patient is required to hold the leg extended with the knee at above a minimum angle for a number of seconds.

With *FriWalk* the exercise is automated enabling the parallel collection of body posture information. The walker is tasked to remind the patient with audio and visual alerts that he must perform the exercise prescribed by the physician and invites him/her to complete it in pre-set intervals. As soon as the patient accepts the invitation, the caregiver moves the walker to the optimal position for the measurement. To achieve that, the system guides the caregiver to the proper position with on-screen visual indicators.

The isometric/isotonic exercise is initiated and the *FriWalk* gives audible and written instructions to the patient. The patient is instructed to lift and hold each leg up a pre-set number of times alternating left and right. During each repetition the patient is required to hold the left up for a given number of seconds.

The display shows two bars with the number of lifts detected, the bars correspond to the left and right leg. During the exercise the counter goes down with each successful repetition indicating the progress to the patient. The counter is not decremented if the patient moves the wrong leg or if the minimum desirable knee extension is not reached. When the knee reaches the right extension a “tick” sound is emitted. A timer is started for the period the patient needs to hold his/her leg up. After the “hold” time has passed a new tick is emitted. If the patient does not succeed in holding the position long enough, the system records the partial failure and the counter is decremented.

During the exercise the *FriWalk* is recording the posture of the patient and the angle of the knee for each leg. This information provided by FORTH’s body tracking module is stored and made available to the physician for later analysis. The body tracking module is also responsible for informing the user interface about the knee angles in real time as the exercise progresses.

Once the number of repetitions is reached the system informs both the patient and the caregiver that the exercise is completed and the *FriWalk* returns to normal operation mode.

Chapter 4

User State Modeling (USM)

4.1 Introduction

The purpose of the User State Model (USM) is to condense and evaluate the emerging plurality of data presented in Chapter 2 in order to deliver semantically meaningful information that allows for automatic activity evaluation and measuring the impact of activities on the user's well-being. We introduce a hierarchical aggregation of sensory information and present the underlying mathematical models for rating the user's well-being on the physiological and emotion level as well as express the level of vigilance, an activity index and a stress level.

Section 4.2 is dedicated to the exact specification of the input quantities of the USM. In section 4.3, the general structure of the USM is presented: The main data flows are shown and a high-level overview of the calculation types and steps are presented. Finally, section 4.4 focuses on the precise description of the mathematical models of the USM (i.e. its sub-models) and the interrelations between the individual quantities.

4.2 Inputs, Interfaces: External Specifications

The principle data interfaces are sketched in Figure 21, the exact specifications for the Fitbit Charge HR wristband, the gait analysis, and the camera on the face can be found in Table 11.

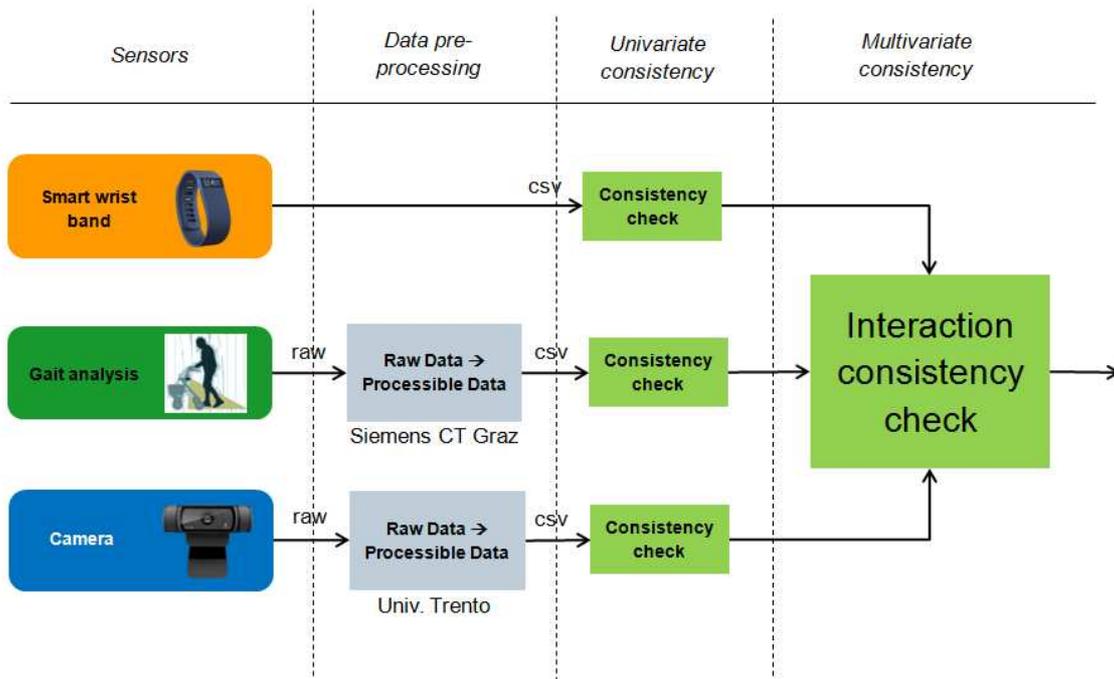


Figure 21: A schematic overview of how the sensor data is handled before the calculation of the User State Model (USM).

| Input (Camera) | | | | | |
|-----------------|----------------------------|------|-----------|-----------|---|
| Parameter | | Unit | Range | Frequency | Comment |
| P | Average: Pain Level | - | [0; 1] | 1min | Pain Level, averaged over 1min |
| maxP | Maximum: Pain Level | | | | Peak value of all Pain Levels of 1min |
| σ_P | SD: Pain Level | | | | Standard deviation of Pain Levels of 1min |
| HR | Average: Heart Rate | bpm | [0; 200] | 1min | Heart Rate, averaged over 1min |
| maxHR | Maximum: Heart Rate | | | | Peak value of all Heart Rates of 1min |
| minHR | Minimum: Heart Rate | | | | Smalles value of all Heart Rates of 1min |
| σ_{HR} | SD: Heart Rate | | | | Standard deviation of Heart Rates of 1min |
| emo | Average: Emotional valence | - | [-1; 1] | 1min | Emotional valence, averaged over 1min |
| σ_{emo} | SD: Emotional valence | | | | Standard deviation of Emotional valences of 1min |
| a | Average: Arousal | - | [0; 1] | 1min | Arousal, averaged over 1min |
| σ_a | SD: Arousal | | | | Standard deviation of Arousal of 1min |
| Yaw | Average: Head Yaw | ° | [-90; 90] | 1min | Averages and SDs as before. If it's possible, you may also hand me just the average SD of all three head poses (instead of the single three SDs). |
| σ_{Yaw} | SD: Head Yaw | | | | |
| Tilt | Average: Head Tilt | ° | [-90; 90] | 1min | |
| σ_{Tilt} | SD: Head Tilt | | | | |
| Roll | Average: Head Roll | ° | [-90; 90] | 1min | |
| σ_{Roll} | SD: Head Roll | | | | |

| Input (Gait analysis) | | | | | |
|-----------------------|--------------------------|------|----------|-----------|--|
| Parameter | | Unit | Range | Frequency | Comment |
| N_{steps} | Number of steps | - | [0; 200] | 1min | Number of steps of 1min |
| SL | Average: Stride Length | m | [0; 5] | 1min | Stride Length, averaged over 1min |
| σ_{SL} | SD: Stride Length | | | | Standard deviation of Stride Length of 1min |
| GCT | Average: Gait Cycle Time | s | [0; 60] | 1min | Gait Cycle Time, averaged over 1min |
| σ_{GCT} | SD: Gait Cycle Time | | | | Standard deviation of Gait Cycle Times of 1min |
| ST | Average: Stance Time | s | [0; 60] | 1min | Stride Length, averaged over 1min |
| σ_{ST} | SD: Stance Time | | | | Standard deviation of Stride Length of 1min |

| Input (Fitbit Wristband) | | | | | |
|--------------------------|--------------------|------|-----------|-----------|-----------------------------------|
| Parameter | | Unit | Range | Frequency | Comment |
| N_{steps} | Number of steps | - | [0; 200] | 1min | Number of steps of 1min |
| cal | Active calories | kcal | [0; 4000] | 1day | Number of burned calories of 1day |
| HR | Heart Rate | bpm | [0; 200] | 1min | Current HR, sent every minute |
| t_{sleep} | Time asleep | min | [0; 500] | 1day | Time asleep at night |
| t_{awake} | Time awake | min | [0; 400] | 1day | Time awake between sleep cycles |
| RHR | Resting Heart Rate | bpm | [0; 120] | 1day | Heart rate right after wake up |

Table 11: Collected input parameters for the User State Model (USM), including their physical units, quantity ranges, and sending frequencies.

4.3 General Structure

A sketch of the general data flow before and after the calculation of the USM output is presented in Figure 22. It shows the data path starting from sensors including pre-processing and USM calculation to sending the USM outputs.

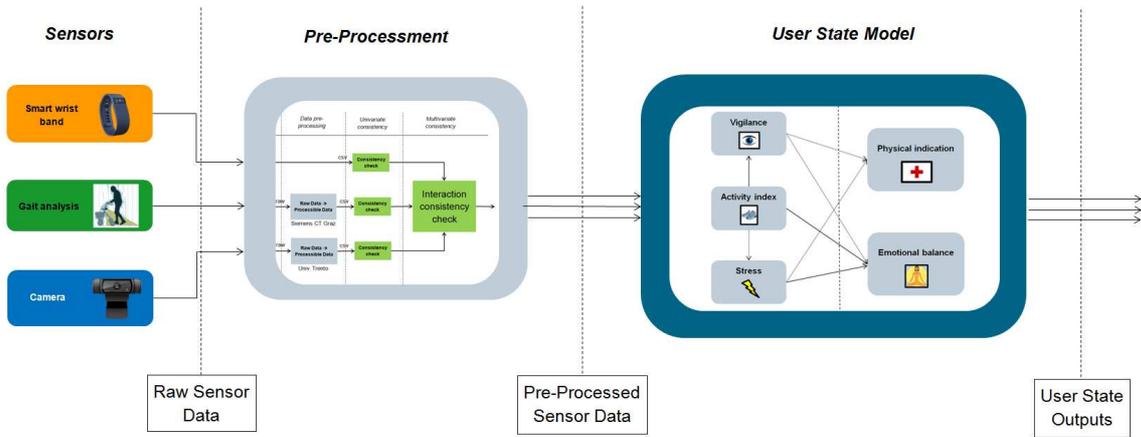


Figure 22: A schematic overview of the data paths before and after calculation of the User State Model (USM); the raw sensor data of the biometric measurements is pre-processed (time-aligned, converted into a uniform data format and checked for consistency), further processed by the USM and sent to a general communication platform

The high-level structure of the USM itself is depicted in Figure 23. “Vigilance”, “Activity index”, and “Stress” are determined as level1 outputs – they are mainly influenced by the sensor measurements and considered hardware oriented or low-level information. “Physical indication” and “Emotional balance” are defined as level2 outputs – they equally take both sensory information and level1 output into account (meta- or high-level information).

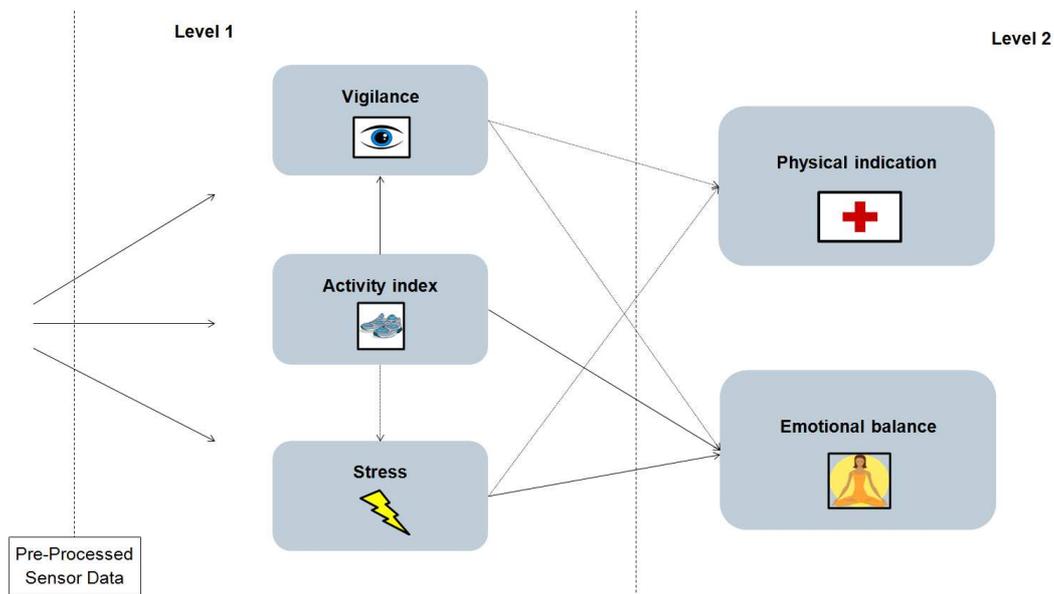


Figure 23: General model architecture of the User State Model (USM), which consists of two levels: the quantities “Vigilance”, “Activity index”, and “Stress” (level 1 outputs) are classified according to their calculation in close vicinity to the pre-processed sensor data; the quantities “Physical indication” and “Emotional balance” (level 2 outputs) are calculated on a meta-level taking both sensor data and level1 outputs into account.

Detailed descriptions of the calculations of each of the level1 and level2 outputs are provided in the following chapter.

4.4 Detailed Structure & Internal Specifications

4.4.1 Model deduction approach

For sake of simplicity, we assume linear approximations of all presented relationships. In future versions of the USM, these relationships may easily be extended to quadratic, logarithmic or any other suitable type of mathematical relationship.

The “range” of every input quantity consists of 4 numbers:

- 1) Absolute minimum: Below that number, the input is not recognized and not considered in the calculation. Above that number, the impact if the input is either maximal or minimal, depending on the slope of the function.
- 2) Linear minimum: At this value, the linear function for the impact applies.
- 3) Linear maximum: Up to this point, the linear function still applies.
- 4) Absolute maximum: Above that number, the input is not recognized and not considered in the calculation. Below that number, the impact if the input is either maximal or minimal, depending on the slope of the function.

See Figure 24 for a graphical representation of the sub-model generation.

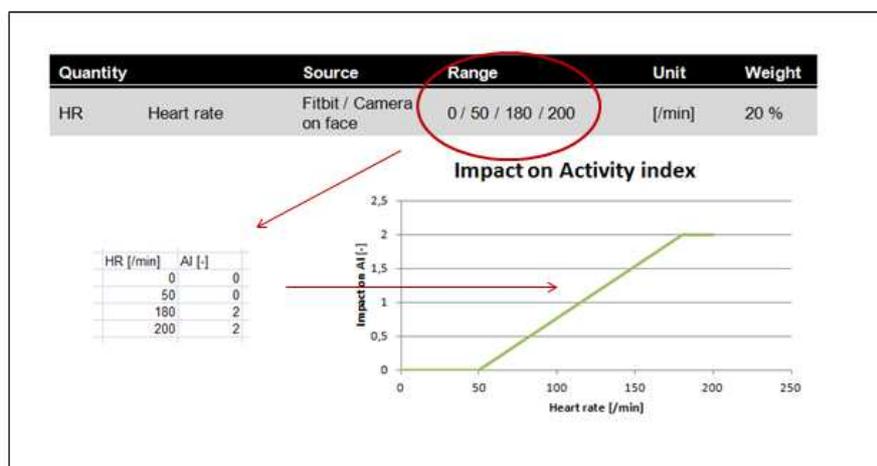


Figure 24: Each quantity range is represented with 4 numbers which determine the linear impact function: the absolute minimum, the linear minimum, the linear maximum, and the absolute maximum.

4.4.2 Activity Index (AI)

Motivation

Activity is strongly correlated to general health and thus should be greatly encouraged: e.g. it prevents cardiovascular diseases [29] and correlates with lower numbers of obesity [30] and general mortality [31]. Due to its key importance, we define the “Activity index” as a crucial field of the ACANTO User State Model.

Inputs

Basing on steps, climbed floors and active versus non-active time, the Fitbit Charge HR gives an estimation on **actively burned calories** per day. As we consider the numbers taken into account as crucial, we weigh this quantity with an impact of 50%.

The **current heart rate** delivers information on the patient’s activity in the very moment of calculation. Although this quantity is measured redundantly (by the camera and the Charge HR), we expect its availability and accuracy to be limited and thus assign a rather low impact of 20%.

Walking speed can be derived from the gait analysis and represents another proper means of measuring a patient’s activity; its impact is set to 20%.

To ensure a broad (and thus robust) calculation of “Activity index”, the **head movements** – measured by the camera on the face – are taken into account as well: the standard deviations of all three spatial dimensions are combined and determine the AI by 10%.

Outputs

The “Activity index” is represented by a decimal number with 0 (not active) and 10 (very active). The objective is to reach an AI of 10. Hyper-activity or other detrimental forms of activity are not considered. The AI is calculated every minute.

Summary and calculation formula

| Quantity | | Source | Range | Unit | Weight |
|-------------------------|---|-------------------------|---------------------|------------|--------|
| <u>cal</u> | Actively burned calories | Fitbit | 0 / 0 / 3000 / 4000 | [kcal/day] | 50 % |
| <u>σ_{Yaw}</u> | Head pose: Standard deviations of Yaw, Tilt, Roll | Camera on face | 0 / 0 / 5 / 7 | [deg] | 10 % |
| <u>σ_{Tilt}</u> | | | | | |
| <u>σ_{Roll}</u> | | | | | |
| <u>HR</u> | Heart rate | Fitbit / Camera on face | 0 / 50 / 180 / 200 | [/min] | 20 % |
| <u>v_{walk}</u> | Walking speed | Gait analysis | 0 / 0 / 2 / 3 | [m/s] | 20 % |

Table 12: Overview of input quantities for Activity index (AI); the quantities, their data source, validity ranges (for their interpretation, see Figure 24), units and impacts for the AI are displayed.

$$AI = \underbrace{0.0017 \cdot cal}_{\text{Calories impact}} + \underbrace{0.0667 \cdot (\sigma_{Yaw} + \sigma_{Tilt} + \sigma_{Roll})}_{\text{Head pose impact}} + \underbrace{0.0154 \cdot HR - 0.7692}_{\text{Heart Rate impact}} + \underbrace{v_{walk}}_{\text{Walking speed}}$$

4.4.3 Vigilance

Motivation

“Vigilance” is another important field for the ACANTO project: An alert client perceives his environment better and suffers from fewer accidents.

Inputs

Amongst other quantities, sleep quality represents an important input quantity for “Vigilance” as it highly affects sleepiness and fatigue [33][70]. This quantity is represented by Fitbit’s **time asleep** and **time awake between sleep cycles**. They are weighed with impacts of 20 % and 10 %, respectively.

As the word “Vigilance” may be used synonymously for **arousal** (being test, alert, excited), this quantity – measured via the camera setup – is a logical input factor. We assume its impact at 40 %.