Ref. Ares(2015)4922931 - 07/11/2015

Health, demographic change and wellbeing

Personalising health and care: Advancing active and healthy ageing

H2020-PHC-19-2014

Research and Innovation Action

ACANTO

A CyberphysicAl social NeTwOrk using robot friends

Deliverable 3.1.1

User perception and platform localization:
First results on holistic user awareness and FriWalk localization

Deliverable due date: 10.2015	Actual submission date: 02.11.2015
Start date of project: February 1, 2015	Duration: 42 months
Lead beneficiary for this deliverable: SIEMENS	Revision: 1.0
Authors: Johannes Scheiblauer (SIEMENS), Pashalis Padeleris (FORTH), Sergey Tulyakov (UNITN), David Macii (UNITN), Daniele Fontanelli (UNITN), Josef Alois Birchbauer (SIEMENS)	
Internal reviewer: Nicu Sebe (UNITN), Antonis Argyros (FORTH)	

The research leading to these results has received funding from the European Union's H2020 Research and Innovation Programme - Societal Challenge 1 (DG CONNECT/H) under grant agreement n°643644

Dissemination Level		
PU	Public	
CO	Confidential, only for members of the consortium (including the Commission Services)	x

Contents

Executive Summary

In this deliverable we present and investigate various **options for sensing the user (biosignals / biometrics**) in a non-obtrusive way by instrumentation of our FriWalk/FriTab and selectively also the user in the case of wearable technology. In this respect we consider commercial off-the-shelf (COTS) devices and we discuss as well our research outcome that bridges the gap in case of missing commercial solutions. We carefully consider desired KPI's of our solutions, e.g. minimum form factors, a reasonable pricing etc. even at the early state of conception.

Next we give a first concept for **modeling the user state** derived from the individual sensor measurements and address the tradeoff between specificity and semantic meaning by introducing a hierarchical aggregation of sensory information.

Localization is a central aspect in ACANTO for giving guidance and destination specific information. Since each FriWalk is a CPS interlinked and being able to communicate with other FriWalk devices we introduce a novel concept of **collaborative localization**.

Chapter 1
Introduction

1.1 Role of this deliverable within WP3

One of the key ideas of ACANTO is to learn as much as possible about the user of the FriWalk *without* the necessity to have this information actively provided by the user, since we want to ease the burden for our target group and not to pose an additional challenge. In other words this means continuous observation and perception of the user's state. Some of the observations will be relevant only at the time of the measurement, some will be meaningful by aggregation over a longer period of time, some of them indicate physiological conditions with medical relevance (e.g. with respect to therapeutic goals) while others address the motivational level or mood of the person. In any case, the means to gather all this information are the sensors. These sensors are deployed on the FriWalk/FriTab or alternatively also on the user, while the latter option is considered very carefully since we want to acquire data in an utmost non-obtrusive way whenever possible. In terms of sensors and also intermediate level information derived from them like biosignals we don't want to reinvent the wheel. Hence it is important to have a look at the available COTS solutions in this domain (in particular on consumer devices which are widespread nowadays due to a general trend to fitness and a healthy lifestyle), which have to satisfy some special requirements in our case. In case such devices or solutions do not exist research has been carried out and preliminary results are presented aiding in rating the feasibility. Condensing the plurality of information into a semantically meaningful user state model is the next step. A tradeoff between specificity and semantic meaning has to be found in this respect, i.e. not all information will be helpful for each recipient. We address this issue by introducing a hierarchical aggregation of sensory information. Platform localization is investigated carefully as well. Since the user is "linked" to the FriWalk we implicitly localize the user via this technique, hence this can also be considered as user sensing. However, the platform localization is different from the perspective that it can use *intrinsic* (e.g. odometry) as well as *extrinsic* (e.g. GPS) cues. Furthermore, localization can also be performed in a *relative* manner, i.e. with respect to other FriWalk units by introducing of a novel concept of *collaborative* localization that reflects an aim of ACANTO in a very natural manner: fostering group activities and social contacts amongst older adults.

1.2 Deliverable objectives

The aim of this deliverable can be summarized as follows:

- Identify suitable solutions for **sensing** of **biosignals and biometrics** in the context of FriWalk/FriTab. This includes commercially available (consumer) devices as well as the validation/proof of concept of scientific approaches in cases a proper COTS solution is lacking according to our knowledge.
- Conception of a **user state model** that is derived from the sensory information
- Develop and present a novel concept for ***collaborative*** or ***synergic*** **localization** of an ensemble of FriWalk units.

1.3 Deliverable organization

The rest of the deliverable is organized as follows: In chapter 2 we present various options on user-centric sensing including wearables, non-contact measurements for heart and respiration rate and higher level person analytics like gait and body analytics. Chapter 3 presents a first concept on how to model and semantically abstract the users state from the plurality of sensory and intermediate level information derived in chapter 2. Of course both chapters are closely interlinked with each other in an iterative sense; the modeling provides us an indication on what sensory information is required (top-down) while the sensor level motivates the processing of existing sensory data (bottom-up). Chapter 4 reports on concepts for localizing the FriWalk (and implicitly the user therefore) in the environment. In particular chapter 4 elaborates on a novel concept for *collaborative* or *synergic* localization.

Chapter 2
User-Centric Sensing (Biosignals / Biomterics)

2.1 Wearables and Smart Fabrics (SIEMENS)

This chapter deals with wearable biometric sensing devices ("wearables") such as wristbands or smart garments. Requirements for these wearables are motivated, results of the subsequent desk search are discussed and results of the corresponding experiments are presented.

2.1.1 Requirements

There are dozens of biometric sensing devices available, but not every wearable is applicable to the ACANTO project. To facilitate the search for appropriate wearables, certain constraints ("KPI's") were defined; they were be classified into the main requirements "Permanent wearability", "Data acquisition and availability", and financial aspects ("Price"). An overview of these requirements and derived sub demands are listed in
Table 1.

MAIN REQUIREMENT	DEMAND	DESIRED VALUE, ANSWER OR RANGE
Permanent wearability	Battery lifetime	5 days and more
	Water proof	Yes
Data acquisition and availability	Steps	Yes
	Sleep	Yes
	Other health related data	Heart rate and beyond
	Synchronization	Bluetooth and via Desktop application
	Data availability	Open system
Price	Price	Below 100€

Table 1: Overview of the requirements and derived demands for the desk search for wearable biometric sensing devices.

Note that the listed demands are suggestions, not all of them are "deal breakers" – if e.g. a wearable is more expensive, but complies with all other requirements, can be still considered a potential sensing device.

2.1.2 Desksearch

According to the definitions in 2.11, an extensive desk search was conducted. See Table 2 for the results.

Manufacturer	Device	Permanent wearability		Data acquisition and availability					Price
		Battery	Water p.	Step	Sleep	Other	Synchro	Availability	Prcie (€)
Garmin	Vivofit2	365	Y	Y	Y	N	BT, Dongle	Open	120
Fitbit	Zip	180	Splash	Y	N	N	BT, Dongle	Open	60
Jawbone	Up move	180	Splash	Y	Y	N	BT	?	60
Misfit	Shine	180	Y	Y	Y	N	BT	Closed	120
a-rival	Quairos	21	Splash	Y	Y	N	BT	?	70
Withings	Heart rate Ox	14	N	Y	Y	HR, Oxy	BT	Open	120
Jawbone	Up24	14	Splash	Y	Y	N	BT	Open	100
Mio	Fuse	7	Y	Y	N	HR	BT	Open	120
Spire	Spire	7	Splash	Y	N	Resp	BT	?	150
Jawbone	**Up3**	**7**	**Splash**	**Y**	**Y**	**HR, T, Persp**	**BT**	**?**	**150**
Medisana	Vifit	7	N	Y	Y	N	BT	Closed	100
Runtastic	Orbit	7	Y	Y	Y	Light, T	BT	Closed	120
Wellograph	Wellness Watch	7	Y	Y	N	HR	BT	?	350
Polar	Loop	6	Y	Y	Y	N	BT	?	70
Huawei	Talkband B1	6	Y	Y	Y	N	BT	?	80
Fitbit	**Charge HR**	**5**	**Splash**	**Y**	**Y**	**HR**	**BT; Dongle**	**Open**	**150**
Basis	Peak	4	Y	Y	Y	HR, Persp, T	BT	Open	200
Basis	Carbon Steel	4	Y	Y	Y	HR, Persp, T	BT	Open	200
Nike	Fuelband SE	4	Splash	Y	N	N	BT	Closed	48
Samsung	GearFit	3	Y	Y	Y	HR	BT	Closed	150
LG	Lifeband Touch	3	Y	Y	N	N	BT	?	70
Sony	Smartband SWR30	3	Splash	Y	Y	N	BT	Closed	140
Hexoskin	**Smart garment**	**3**	**Partly**	**Y**	**Y**	**HR, HRV, Resp**	**BT, Dongle**	**Open!**	**400**
Heart	rateeOn	3	Splash	N	N	HR	BT	Closed	200
Amiigo	Amiigo	2	Y	Y	Y	HR, HRV, Oxy, Resp, T	BT	?	180
Microsoft	Band	2	Splash	Y	Y	HR, T, UV, Galvanic	BT	Open	200

Table 2: Overview of the investigated devices according to the requirements defined in Table 1; battery life (according to which this table is sorted) is given in days, *"Y/N"* stand for Yes/No, *"HR"* is heart rate, *"HRV"* is heart rate variability, *"Oxy"* is blood oxygen, *"Resp"* is respiration, *"T"* for temperature, *"Persp"* for perspiration, *"UV"* for ultra-violet light detection, *"BT"* stands for Bluetooth; green shades cells indicate satisfied demands, bold-faced rows represent chosen devices.

Table 2 is sorted by decreasing battery life (in days); it can be seen that long-living devices (> 150 days) lack functionality in terms of other medical measurements as e.g. heart rate. The "*Withings Heartrate Ox*", "*Jawbone Up24*", "*Mio Fuse*", and "*Spire*" are either not water proof, lack additional functions or do not support sleep tracking.

The "***Jawbone UP3***" and the "***Fitbit Charge HR***" are similar products, with the former being a promising new device and the latter being an established wearable device well rated by other costumers. Both products are seen suitable for the testing phase of the ACANTO project. Other products were withdrawn from the list of potential wearables due to their lack of sleep tracking or other medical measurements, short battery time or their sensitivity to water.

Despite the relatively high price, the "***Hexoskin Smart garment***" is a promising wearable. It is well rated and offers a lot of additional measurements as well as it allows a complete access to each measured quantity. Its short battery life might impede its permanent use, but it could serve as an appropriate ground truth for the medical measurements.

Due to their (in comparison) very high prices and/or very low battery life, smart watches like e.g. the "*Apple Watch*" or the "*Motorola Moto 360*" were not considered in Table 2.

2.1.3 Experimental Results

In this section a review of the experiments with the chosen devices is presented. A check on the promised features of Table 2 is performed and further remarks are made. At the end of this section, the wearable devices are compared.

"Jawbone Up3"

Battery lifetime	The battery of the "Jawbone Up3" lasts about 5-6 days, which is below the promised 7 days but still close enough.
Water proof	Accidental splashes of water (e.g. while washing dishes) did not harm the device; further tests regarding its water resistance have not been conducted.
Steps	Steps are counted in a conservative manner: accidental shocks or clapping are not recognized as steps, but also a fraction of real steps is missed. See Figure 2 (left) for a screenshot of the activity screen of the smart phone app.
Sleep	Sleep is not recognized completely independent, but in a detailed manner: Awake phases as well as phases of light, deep and REM sleep are distinguished (see the right picture of Figure 2).
Other health data	The only non-derived medical quantity is the resting heart rate, which is measured once a day before wake-up. The promised values for temperature and perspiration are either not measured or not displayed.
Synchronization	The "Jawbone Up3" delivers every of its 48 measured quantities as a daily average. Communication to the „Jawbone" server is only possible via Bluetooth, thus a new-generation smart phone is mandatory.
Data availability	Data is available in CSV format after login into the „Jawbone" home page.
Price	The price before taxes (€172) was – due to supplier difficulties – higher than assumed.
Other remarks	The device has no display; the only communication with the user is via vibration. Its wearability is limited due to a poorly designed clip which leads to accidental opening and potential loss of the device.

Table 3: Results for the "Jawbone Up3" after first tests, considering the requirements defined in Table 1 and regarding the promised features depicted in Table 2.

Figure 1: Commercial picture for the “Jawbone Up3” tracker.

Figure 2: Screenshots of the “Jawbone” smart phone app,
depicting activities of one day (left) and sleep analysis (right).

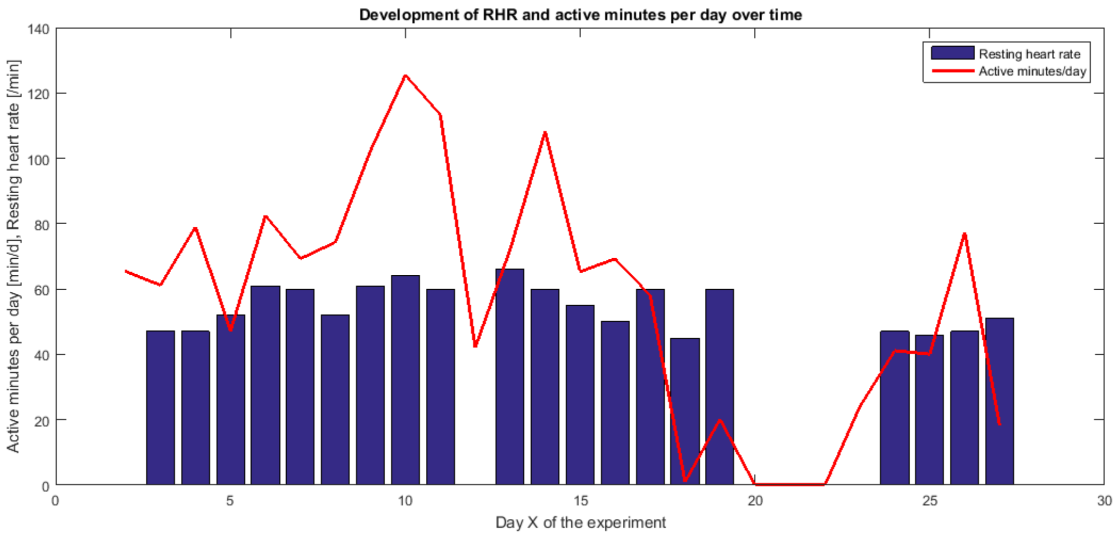


Figure 3: Possible representation of „Jawbone“ data: Resting heart rate (blue bars) and active minutes per day (red line) over time; zero values indicate in both plots indicate an interruption of the experiment by the user, not malfunction of the device.

“Fitbit Charge HR”

Battery lifetime	The battery of the “Fitbit Charge HR” lasts about 4-5 days, which hits the promised 5 days.
Water proof	Accidental splashes of water (e.g. while washing dishes) did not harm the device; further tests regarding its water resistance have not been conducted.
Steps	Steps are counted in an offensive manner: no steps are missed, but accidental shocks or clapping might be recognized as steps. See Figure 5(left) for a screenshot of the activity screen of the smart phone app. Note that not only steps, but also climbed floors are accurately recognized.
Sleep	Sleep is recognized completely independent in a superficial manner: Awakeness as well as phases of restlessness are distinguished (see the right picture of Figure 5).
Other health data	The only non-derived medical quantity is the heart rate, which is measured continuously via the optical “PurePulse” technology. The resting heart rate is determined via a smart averaging algorithm.
Synchronization	The “Fitbit Charge HR” delivers every of its 17 measured quantities as a daily average. Communication to the „Fitbit“ server is possible via Bluetooth (and a new-generation smart phone), but also via an USB dongle. Unfortunately, some data are not openly available.
Data availability	This data is available in CSV format after login into the „Fitbit“ home page.
Price	The price before taxes (€110) was – due to supplier discount – lower than assumed.
Other remarks	The device has a display with all key quantities displayed. Its wearability is great due to a well designed clip.

Table 4: Results for the “Fitbit Charge HR” after first tests, considering the requirements defined in Table 1 and regarding the promised features depicted in Table 2.

Figure 4: Commercial picture for the “Fitbit Charge HR”.

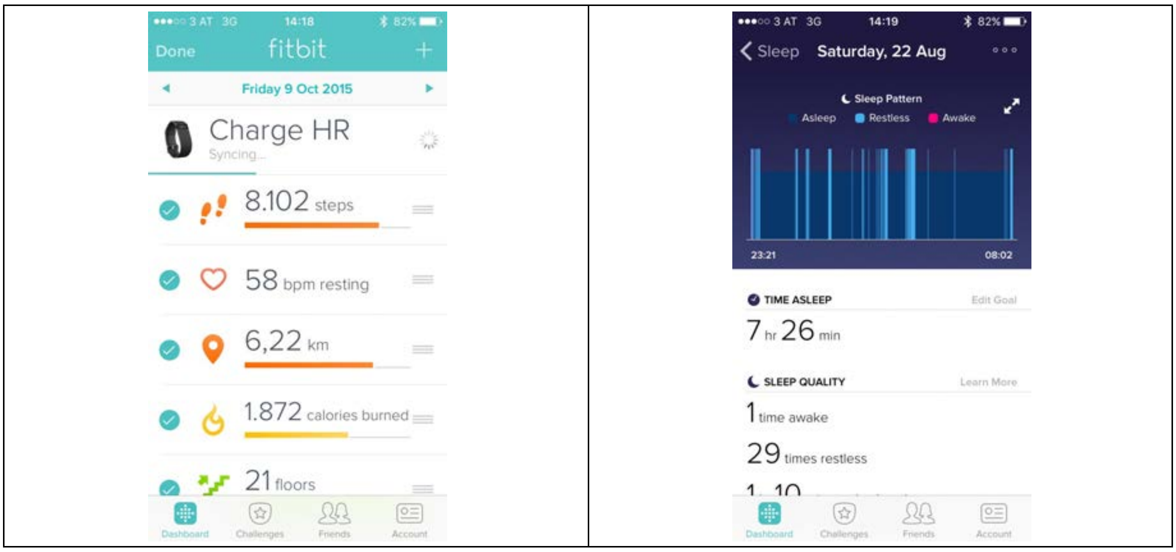


Figure 5: Screenshots of the "Fitbit" smart phone app, depicting activities of one day (left) and sleep analysis (right).

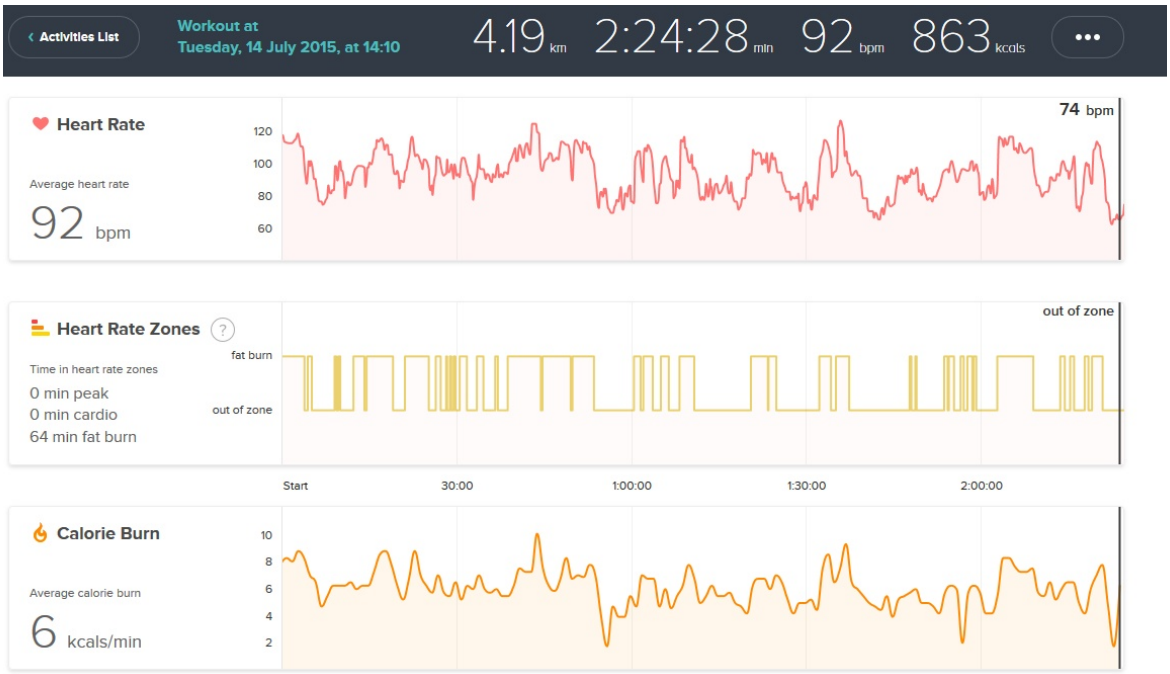


Figure 6: Analysis chart of a "Fitbit" workout, displaying the continuously measured heart rate, the according heart rate zones and calorie burns.

“Hexoskin” smart shirt

Battery lifetime	Since the “Hexoskin” was only used as a reference, no long-time experiments for battery life determination have been conducted; the short-time experiments suggest that the promised 14 hours of intensive recording are a good estimation.
Water proof	The shirt itself is machine-washable, the recording device is endures splash water; since showering with the „Hexoskin“ smart shirt does not make sense, this point is of minor importance.
Steps	Steps are counted in a conservative manner: accidental shocks or clapping are not recognized as steps, but also a fraction of real steps is missed.
Sleep	Experiments for sleep analysis have not been conducted yet.
Other health data	The device measures the patient’s real time ECG and acceleration in all 3 spatial dimensions; these data are used to derive heart rate, heart rate variability, breathing rate, breathing volume, steps and other.
Synchronization	The „Hexoskin“ smart shirt delivers each point of measurement of its 21 quantities. Communication to the „Fitbit“ server is possible via Bluetooth (and a new-generation smart phone), but also via an USB dongle.
Data availability	All data is openly available. Due to the high measurement frequency, about 100 MB of data are produced during one hour of measurement. Basic data is available as CSV, detailed raw data is accessible as binary data (which can be converted to CSV).
Price	The price before taxes (€549) was – due to supplier difficulties – higher than assumed.
Other remarks	The shirt delivers accurate measurements and can be seen as gold standard for the measurements of heart rate and breathing rate. Due to its battery life, the price and the wearability – the shirt fits very tightly – it is not suitable for broad field experiments, but greatly applicable in the validation of other biometric sensors.

Table 5: Results for the “Hexoskin full kit” after first tests, considering the requirements defined in Table 1 and regarding the promised features depicted in Table 2.

Figure 7: Commercial picture for the “Hexoskin full kit”, including the smart shirt (left), the recording device (middle), and the proprietary charging/uploading cable (right).

Figure 8: Screenshots of the Hexoskin analysis dashboard, with the basic metrics (heart rate, breathing rate, minute ventilation, activity, and cadence) on the left and the raw data (ECG, respiration, and acceleration in all 3 dimensions) on the right; note the different time scales.

Overview:

The key findings are summarized in Table 6.

Device	**Battery**	**Water**	**Steps**	**Sleep**	**Other**	**Synchro**	**Open**	**Price**
Jawbone	+	0	+	++	–	–	+	–
Fitbit	0	0	+	+	0	+	0	+
Hexoskin	–	0	++	?	++	++	++	– –

Table 6: Overview over fulfillment of the requirements defined in

Table **1** after first test runs, with the used symbols standing for the following:
“– –“ … *“Not fulfilling the requirements”*
“–“ … *“Barely fulfilling the requirements”*
“0” … *“Acceptable”*
“+” … *“Good”*
“++” … *“Requirements exceeded”*

2.2 Smart Shoe Insoles (SIEMENS)

2.2.1 Requirements

In contrast to the wearable devices presented in the previous section – products which have been developed merchandised and improved for years now – the market for smart shoe inlays is still developing. Since only few products are already commercially available, the requirements for shoe inlays are less rigid than for wearables as given in

Table 1. Note that the shoe inlay requirements presented in Table 7 are again not hard constraints but suggestions.

MAIN REQUIREMENT	DEMAND	DESIRED VALUE, ANSWER OR RANGE
Permanent wearability	Battery lifetime	1 day and more
Data acquisition and availability	Pressure	Yes
	Temperature	Yes
	Steps	Yes
	Synchronization	Bluetooth (or similar)
	Data availability	Open system
Commercial availability	Commercial availability	Yes

Table 7: Overview of requirements & derived demands for the desk search on smart shoe inlays.

2.2.2 Desk search

As mentioned before, the market of smart shoe inlays is still at the very beginning. See Table 8 for the results of the desk search of the shoe inlays regarding the requirements defined in Table 7.

Device	Battery	Pressure	Temp.	Steps	Synchro	Data	Availab.
Moticon OpenGo Science	**14**	**Y**	**N**	**Y**	**ANT+, USB**	**Y**	**Y**
Lechal Insole	3	N	N	Y	BT	Y?	Y
OpenShoe	0.06	Y	N	?	BT	Y	N
Retisense Stridalizer	7	Y	N	Y	BT	N?	N
Digitsole	3	N	Y	Y	BT	?	Y
VitaliShoe	?	Y	N	Y	?	?	N
Uniklinik Magdeburg	?	Y	Y	?	?	?	N

Table 8: Overview of the investigated shoe inlays according to the requirements defined in Table 1 battery is given in days, *"Y/N"* stand for Yes/No, *"BT"* stands for Bluetooth; green shades cells indicate satisfied demands, bold-faced rows represent chosen devices.

Only the “OpenGo Science” device from “Moticon” arises as a potential device for ACANTO. Besides the commercial availability, it offers high-resolution pressure measurement while offering various communication channels (either via ANT+, a network protocol, or via USB) and open data.

2.2.3 Experimental Results

Battery lifetime	Battery lifetime is not a limiting factor, since the button cells last about two weeks; the memory capacity of the insole is – at a recording rate of 50 Hz – limited to about 6 hours.
Pressure	Each insole of the “Moticon OpenGo Science” is equipped with 13 capacitive pressure sensors (see Figure 9).
Temperature	The influence of temperature is neglected.
Steps	Since each step is analyzed in detail, steps are counted at maximum precision.
Synchronization	For financial reasons, we opted for the ANT+ communication only. For every minute recorded (at 50 Hz), the data transfer will last about 30 seconds.
Data availability	All data is available in CSV format.
Commercial availability	The product is only available directly from the producer, starting at €4.000 (before taxes) for a basic version.
Other remarks	In Figure 10, you see a screenshot of the “Moticon Beaker 5”, the associated software package, which allows the collection and analysis of the data. There are two different recording modes: While “Live Capture”, the current pressure distribution is visible and the insoles have to be connected to the computer all the time (via ANT+ stick); while “Recording” no live information is visible (the data is saved to the insoles), but no connection to the computer is necessary. The “Analyze” tab in the “Beaker 5” software package allows a broad analysis of all used pressure sensors, including several derived quantities such as centre of pressure (indicated in black cross in Figure 10).

Table 9: Results for the “Moticon OpenGo Science” shoe inlays after first tests, considering the requirements defined in Table 7 and regarding the promised features depicted in Table 8.

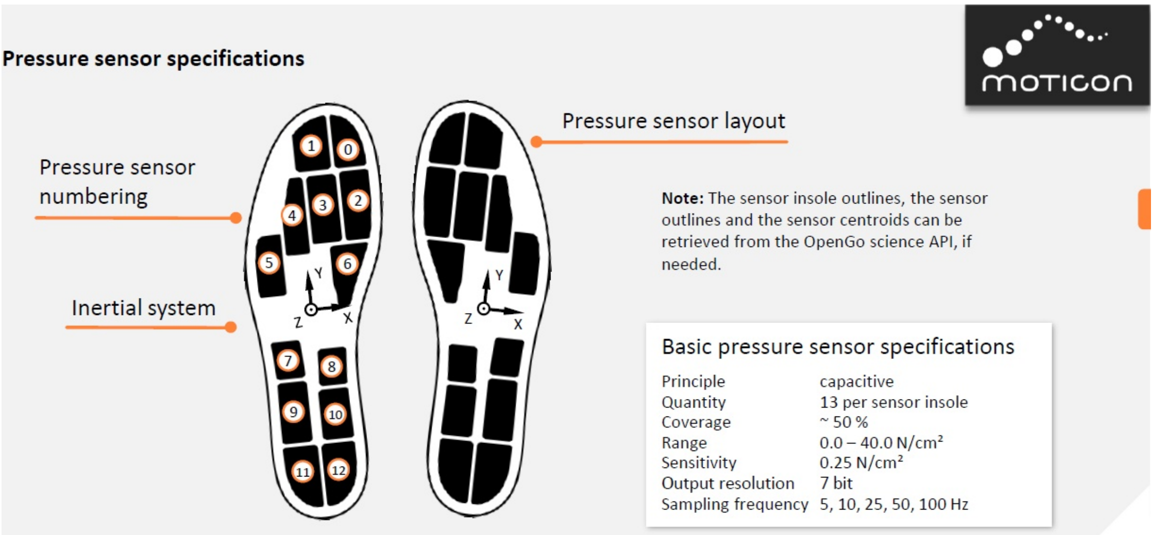


Figure 9: Part of the "Moticon OpenGo Science" specification, depicting the arrangement of the pressure sensors.

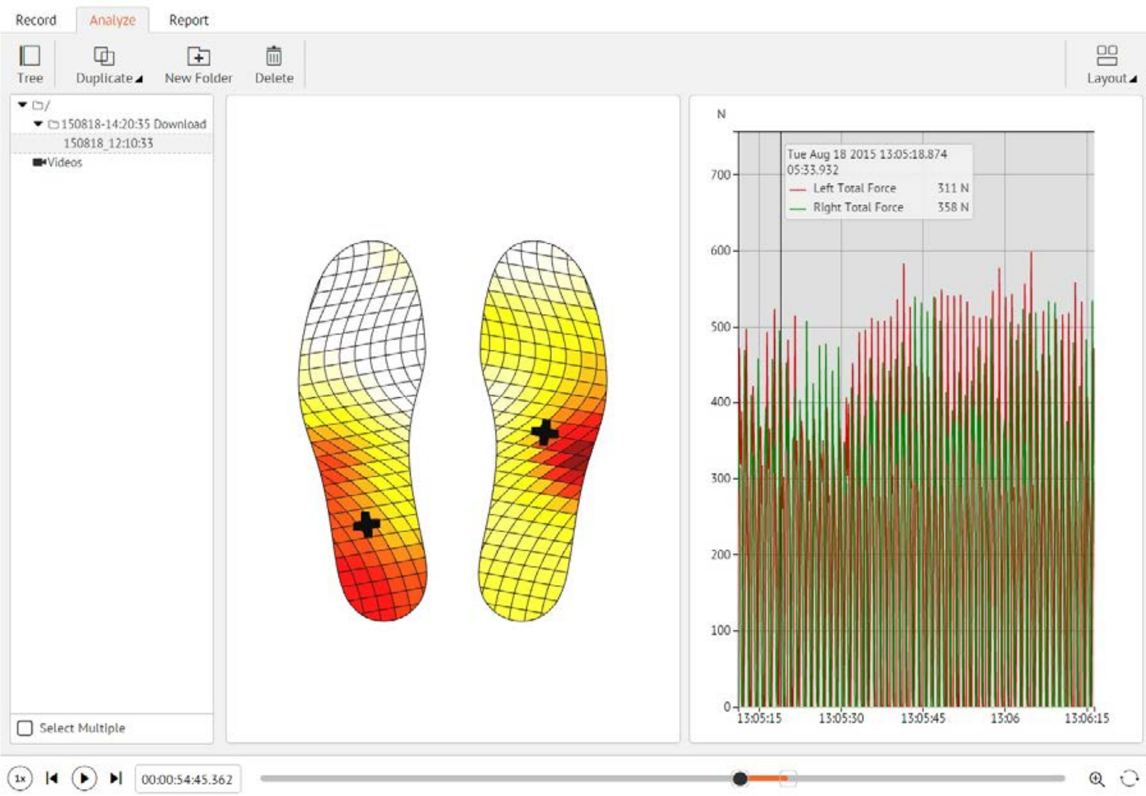


Figure 10: Screenshot of the *Moticon Beaker 5*,the software package for measurement recording and analysis; depiction of the tree structure of experiment protocol (left), pressure distribution of the shoe insoles at a certain point in time (middle), and the total force per foot over time (right).

2.3 Non-contact Respiration Pattern Measurement (SIEMENS)

Besides measuring the pulse (i.e. heart rate) the respiration rate and respiration pattern can reveal important details on peoples current and general physical condition, e.g. if and you much the respiration pattern is reacting when walking on an ascending slope. During the desk search we identified an innovative solution that claims being capable of reading the respiration pattern from distance via radar measurement [1]. The website of the manufacturer explicitly mentions "**senior health care**" as a potential field of applications: *"There are also many opportunities to use XeThru technology for home-based senior health care. While maintaining senior's privacy, the XeThru respiration module can monitor their respiration throughout the night and alert caring family members of any pattern irregularities. Seniors can be given the opportunity to remain independent and live longer in their homes, while family members are granted peace of mind knowing they can still provide care for their loved ones even if they are living far away"*

A white paper available from [1] describes the technical details. The key principle is to measure in sub-mm resolution distances variations in the chest movement.

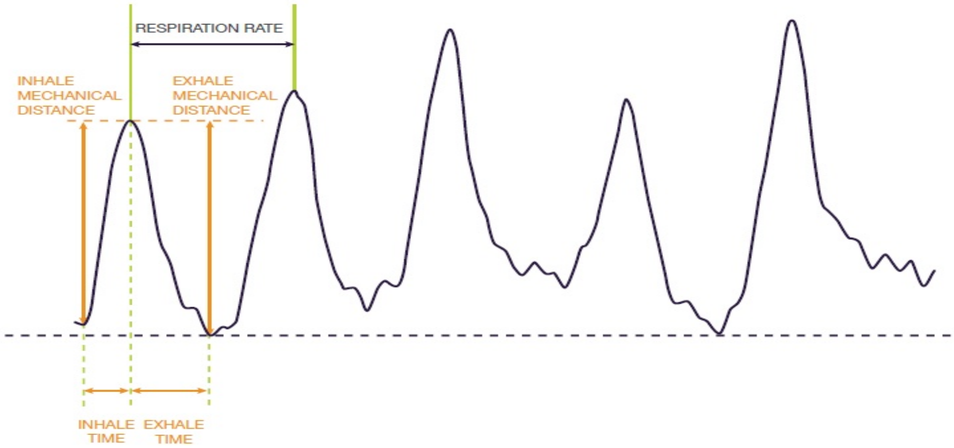


Figure 11: Respiration pattern from adult sitting at range 2 m (taken from [1]).

The following figure shows the components of the sensing module (disassembled for better visibility), the main PCB measures around 6 cm in width. The device uses Ultra-Wideband (UWB) Impulse Radar, and is *"…able to provide medical grade data in a consumer setting, at a low cost."* according to manufacturer information.

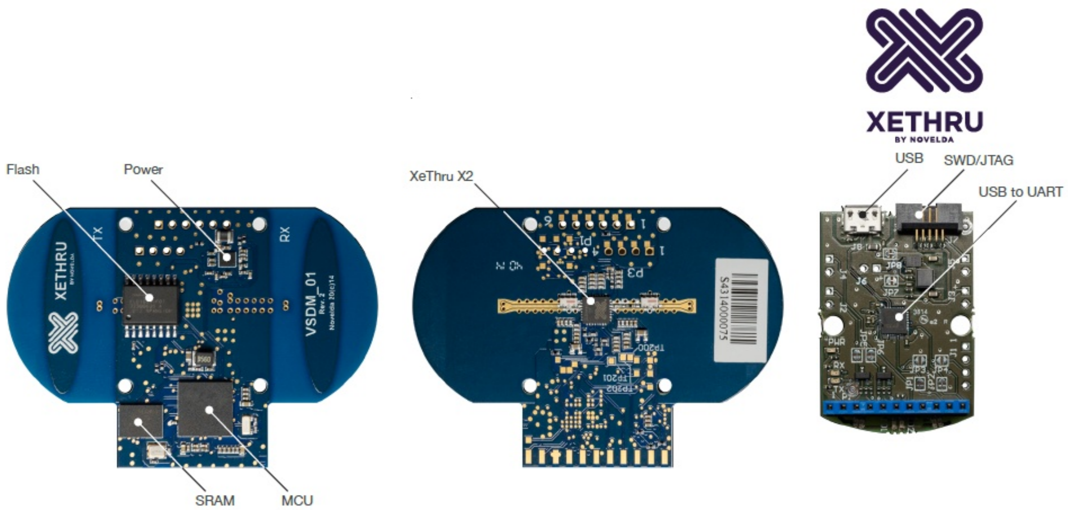


Figure 12: XeThru module and interface card.

In the white paper there is already a factor mentioned that will limit our practical application in the scope of ACANTO: *"Respiration rate is the number of breaths a person takes per minute, and is best measured when a person is at rest."* The key advantage of the technology is being non-invasive so it does not need to be attached to the chest, but could be installed on the walker, which of course violates the constraint of being at rest. According to the manufacture it works highly accurate for distances up to 5 m. Hence we performed a series of practical experiments so verify how a person in move would alter the measurement performance. We use the smart garment **Hexoskin** for that purpose.

Figure 13 shows our des setup on desk as well as on the walker.

Figure 13: Test setup on desk (left) and on the walker (right).

Both experiments revealed a relatively poor performance of the XeThru module. We have been in contact with the manufacturer several times to identify a potential problem on our side but everything seems to be set up properly. We even tested two different modules to exclude the likelihood of a defect one.

The following diagrams compares the results from the XeThru module compared to Hexoskin for a short ride with the walker. The deviation is clearly observable. Not only is the XeThru module incapable of measuring a respiration rate most of the time but also the positively reported values show a significant variation. This is in contrast to what we observe from the Hexoskin which appears much more plausible for a short ride on a horizontal ground by a young adult. Hence we have to conclude XeThru is not useful in the scope of our ACANTO project at least in the current firmware/technology version.

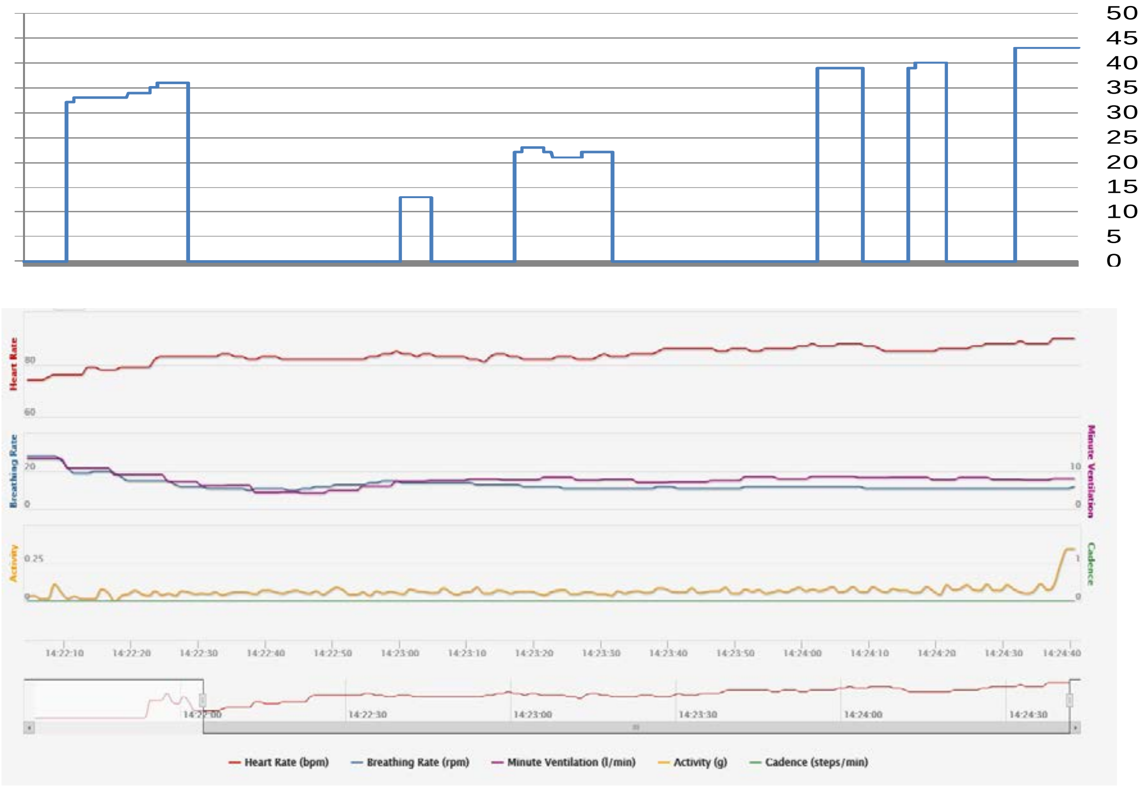


Figure 14: Respiration pattern on a short ride with the walker. XeThru (top) vs. Hexoskin (bottom). The relevant figure ("breathing/respiration rate") is the **blue diagram** in both cases. XeThru delivered inconsistent values during this short ride while Hexoskin tracked more plausible values.

2.4 Non-contact Pulse Measurements (UNITN)

The possibility to detect subtle changes in videos of human skin has been recently investigated and has attracted a lot of attention in the community. These changes appear both in skin color and in subtle motions, and are caused by internal functioning of the heart (Figure 15). Since faces appear more frequently in videos, and due to significant improvements in face alignment methods, many research groups have tackled face-based remote heart rate analysis.

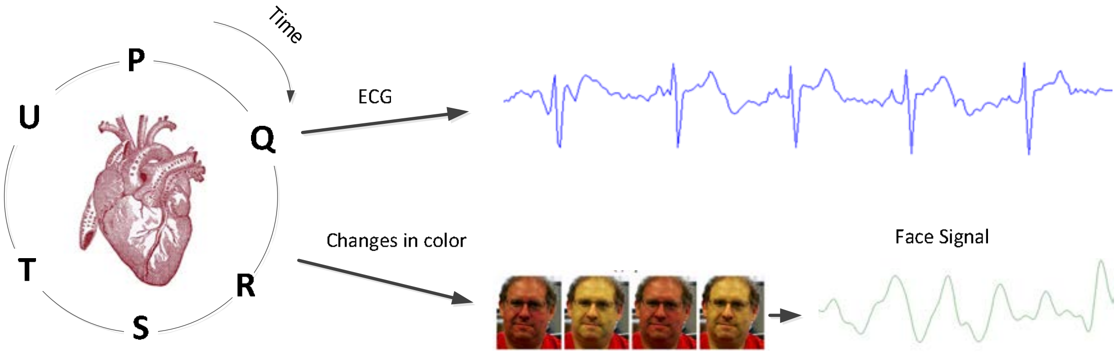


Figure 15: The heart circle consists of the following states: P Q R S T U. These states can be measured by an ECG device by connecting several leads to the subject's body. As it was recently showed, several of those states can be detected via Remote Heart Rate Measurement.

Our pipeline for remote heart rate measurement involves several stages of processing (Figure 16):

1. **Face registration.** It was shown that face registration has significant influence on forthcoming steps, since if done poorly it adds high frequency noise, which is difficult to filter afterwards.

2. **Face region extraction.** It is preferable to select a region less affected by movements and facial expressions.

3. **Signal processing 1.** During the first stage of signal processing the signal is filtered with digital filters.

4. **Signal processing 2.** During the second stage of signal processing, power spectral density estimation methods are used to determine the heart rate

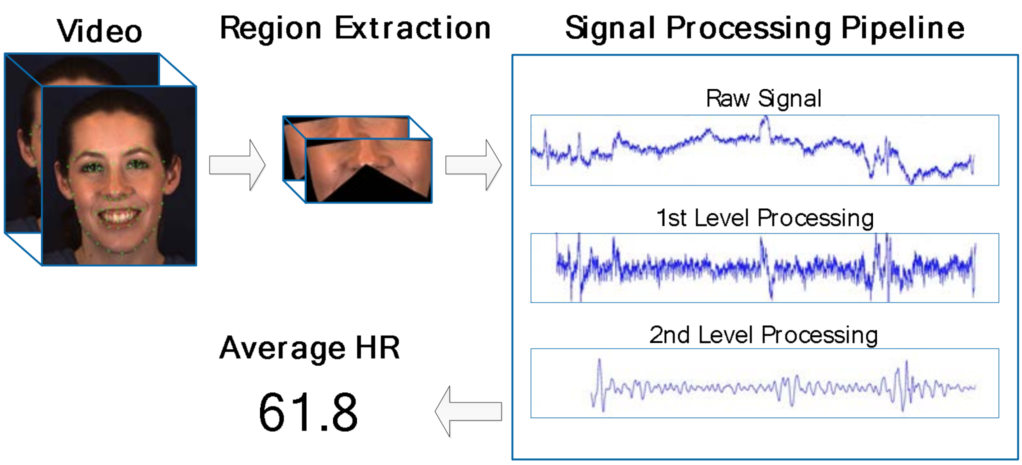


Figure 16: Pipeline for remote heart rate measurement.

In the ideal situation, when there are no motions (see Figure 17) and facial expression, many of the available methods in the literature [2][3][4][5][6][7] can reliably measure the heart rate. However, despite many attempts made, remote heart rate analysis still suffers from major limitations. Most of the methods impose constraints on subject's movements requiring the absence of facial expressions and mimics. These constraints do not hold in real world scenarios envisioned in ACANTO. Clearly, those methods are highly affected by the increased level of noise that appears when subjects behave naturally.

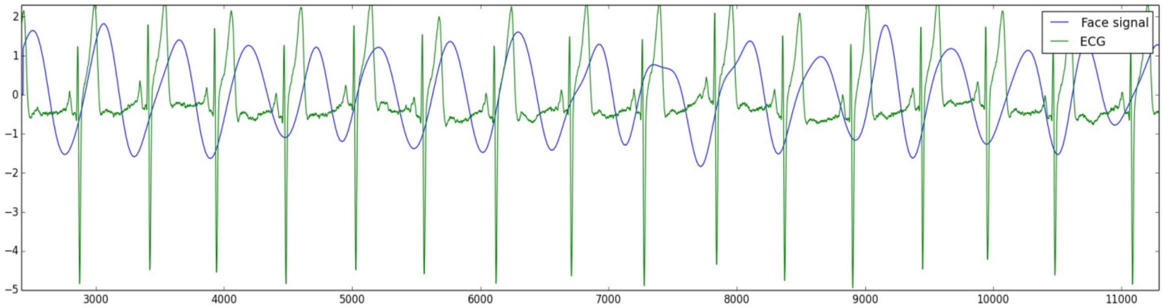


Figure 17: The green signal shows the ground truth ECG, the blue signal shows the output of [2]. The subject did not move during recordings.

Another major limitation is the long time-window analysis. This limitation transforms instantaneous heart rate measurement into average heart rate measurement over a large video sequence. The larger the minimum supported sequence length is the worse is it for real-world applications, since interesting short-time phenomena, such as sudden heart rate increase/decrease due to some emotion, could be missed by such analysis.

2.4.1 Heart rate features

In contrast to common computer vision methods, where the features extraction step (e.g. SIFT) is followed by a problem-specific machine learning approach, a typical pipeline for remote heart rate measurement includes handcrafted filtering methods. This is because it is very difficult to define what a good heart-rate feature is.

Usually one of the most desired property of features is some type of invariance, be it color invariance or scale invariance. However, in case of heart rate features, due to the nature of the signal, such invariance is not desired. Therefore, a different type of features is required. To tackle this problem we introduce what we call heart rate features. These features are invariant to movements and facial expression, while being sensitive to changes in color caused by heart rate activity.

2.4.2 Heart rate sensors

To tackle the aforementioned difficulties, we treat a face as a set of heart rate sensor where each sensor conveys the same underlined signal. The benefit of this idea is that each sensor will capture the noise caused by motions in a different way (Figure 18). Our preliminary studies have shown that this approach is able to deal with the noise induced by head motions and the facial expressions. The current plan is to investigate the robustness of the proposed methodology in the application context of ACANTO.

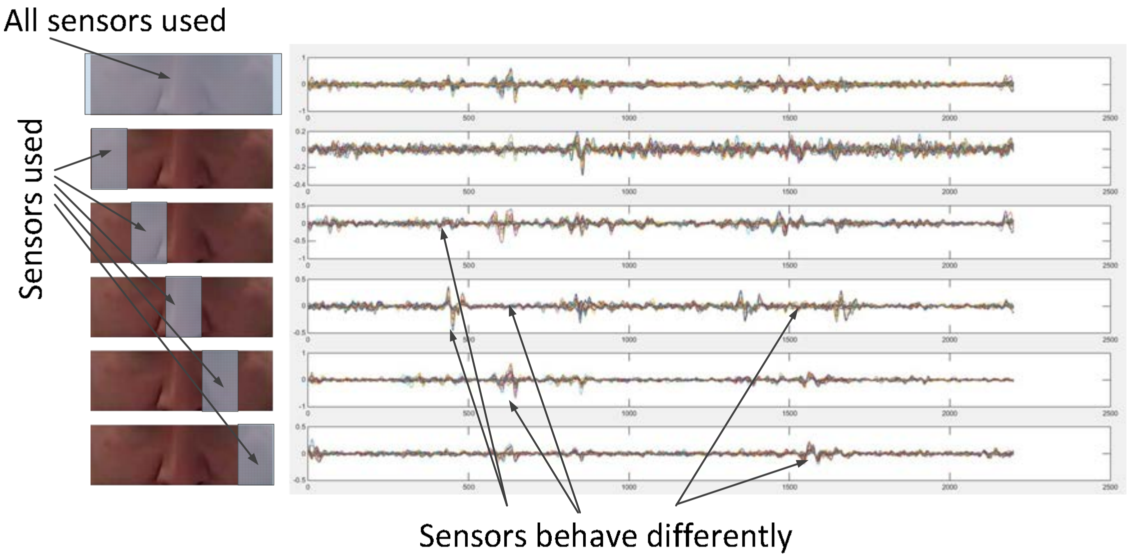


Figure 18: Subject's motions affect different sensors differently. Therefore, we are able to get a clean signal by suppressing the contaminated sensors.

2.5 Depth Cameras and Imaging Sensors (SIEMENS)

Perception in 3D is essential to the ACANTO project, both for the environment as well as for acquiring information about the user. Since a lot of novel sensors/devices are entering the market we did rate them for applicability to the FriWalk and evaluate their performance in various experiments. Details on this evaluation as well as a suggestion on the placement of vision sensors on the walker are given in D3.2.1 *"Perception of the Environment (preliminary): Preliminary results on visual FriWalk sensing system"* which is released simultaneously with D 3.1.1.

2.6 Gait Analysis (SIEMENS)

In order to analyze human gait patterns, highly accurate data must be collected at high frame rates. The state of the art is to deploy a carpet-like structure instrumented with pressure sensors, which allows for measuring position, orientation and pressure of each foot at each step. Progress in new technologies has given rise to devices and techniques that allow for objective evaluation of various gait parameters, resulting in more efficient measurement and providing specialists with a large amount of reliable information on patients' gaits. This reduces the margin of error caused by subjective techniques. Two such measurement tools commonly used in clinical gait evaluation are force platforms or gait walkways, the latter being a carpet like structure instrumented with pressure sensitive elements (sensels). One system that is now in common use is the 'GAITRite®' [10][11].

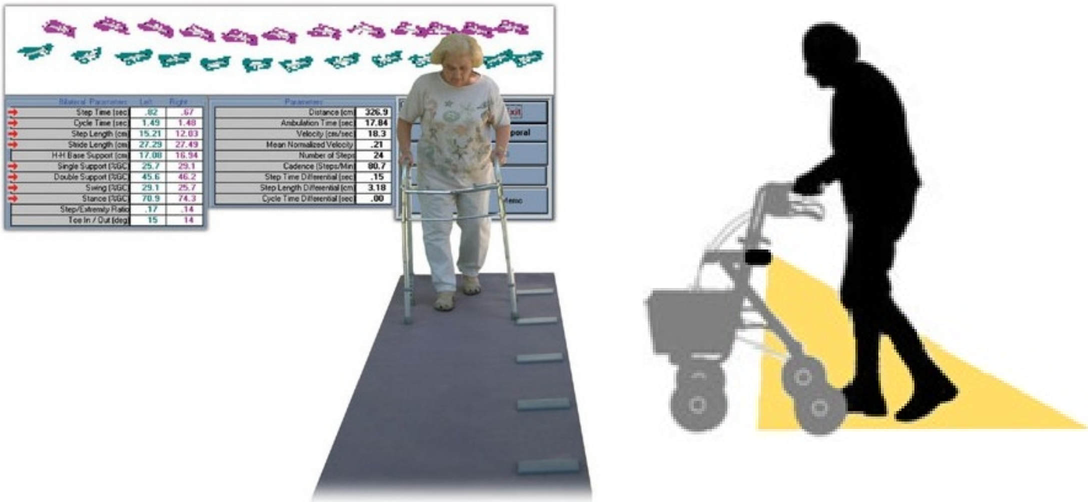


Figure 19: GAITRite® instrumented walkway system (left) and a schematic view of our instrumented walker (right)

Since such gait "walkway carpets" are highly expensive[1] and also limited in length, we propose an alternative in the form of a wheeled walker equipped with a consumer depth camera. We have designed and implemented algorithms that derive the same set of parameters from the depth data as in a gait walkway system, however without the need for the physical presence of a walkway carpet. Moreover, we are able to provide additional information, due to continuous observation of the gait cycle, i.e. not only when the user steps on the ground. In order to retrieve actual foot pressure information, we use a shoe insole sensor (i.e. Moticon).

A detailed technical description can be found in the **Annex** of this deliverable in a paper that has been accepted for presentation at the TechAAL 2015. In addition prior to the publication an invention report has been submitted.

2.6.1 Experimental results

Typical output produced by our system is shown in the subsequent figures. Figure 20 shows that the same data is generated as in the physical gait walkway in Figure 19, i.e. the feet's position, orientation and pressure distribution at each step. While the figure only shows a short sequence, every step the user takes is visualized and the data is stored to disk for further analysis. Figure 21 shows a sample trajectory of the foot tips. It illustrates how our

[1] [1] According to a desk search on various vendors between 25k€-50k€.

system is not only capable of generating data at each step on the ground, but also during the swing phase.

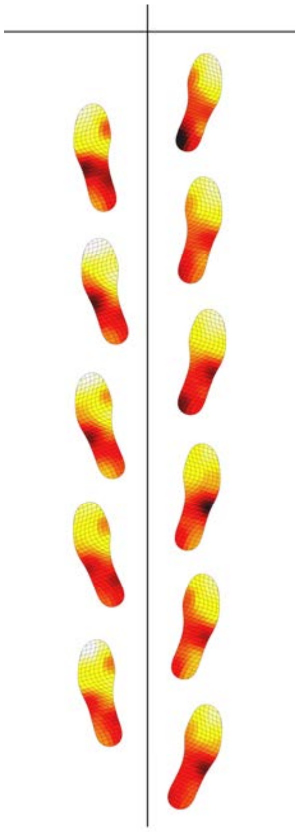


Figure 20: "Virtual Walkway" result sample. The Moticon pressure information is visualized for each step in colors

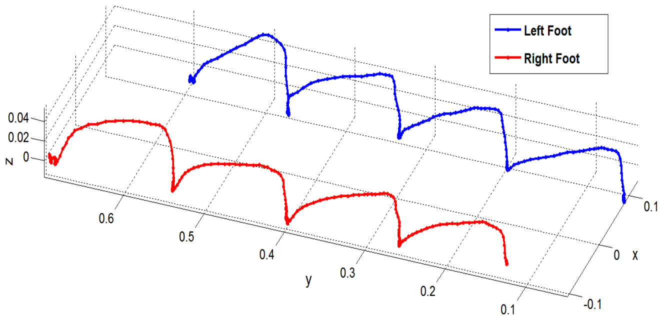


Figure 21: Sample Trajectory of the Foot Tips

The algorithms in Section 5 are designed with a strong focus on speed, which makes it possible to achieve the desired frame rate of 15-20 Hz on a single Intel®-i7 CPU core using a depth map resolution of 640x480 pixels. If higher frame rates are required, the depth map can be sub sampled to around a quarter of the resolution without influencing the results, making frame rates at around 30 Hz possible.

In order to estimate the accuracy of both position and orientation, we performed an extensive evaluation using the Microsoft® Kinect™ sensor. For ground truth generation, we printed several identical feet patterns and placed them at different positions and angles behind the walker. Since absolute trajectories and positions are not relevant for gait analysis, but only the accuracy at each single step matters, we measure relative angles between the patterns and the distances between the foot tips.

As shown in Table 10, the average position accuracy turned out to be slightly less than 3 mm, evaluated in 20 measurements. The error is independent of the step length. Part of the deviation can be explained by the average 3D point resolution of ~1.5 mm and minor inaccuracies at ground truth capturing. Table 11 shows the results of the angle accuracy evaluation. The error increases with the angle, mainly due to occlusions. However, at typical angles when walking (0-15°) the average error of 1.6° is only slightly higher than the ground truth accuracy.

N	μ_{error}	Med_{error}	σ_{error}
20	2,96 mm	2,93 mm	1,68 mm

Table 10: Position Accuracy

Angle	N	μ_{error}	Med_{error}	σ_{error}
0°–15°	40	1,62°	1,39°	1,17°
15°-30°	40	2,26°	1,86°	2.05°
30°-45°	40	3,16°	2,86°	2,18°

Table 11: Angle Accuracy

For comparison, we have evaluated the accuracy of an inertial measurement unit (IMU), namely the Inertial Elements Osmium MIMU22BT [12]. Osmium produces MIMUs (multi IMU) that operate by fusing the measurements of several low cost sensors resulting in enhanced measurement performance. The Osmium MIMU22BT is closely related to the OpenShoe project, an open source foot-mounted inertial navigation system (INS) [13] initiative. We used OpenShoe scripts for data acquisition. While manual calibration can be performed for each individual device using a special calibration object, we used the manufacturer default calibration for practical considerations regarding a potential later deployment, i.e. for being applicable for our target group simplicity in deployment is a factor of high importance.

As shown in Figure 14, the IMU has been attached to the tip of the foot. Table 12 shows the evaluation results. Compared to our results, it turns out that the angles can be measured more accurately using the IMU, but the position error is significantly higher.

Figure 22: IMU attached to the tip of the foot.

μPosition	MedPosition	σPosition	μAngle	MedAngle	σAngle
7,3 mm	6,0 mm	6,6 mm	0,65°	0,50°	0,50°

Table 12: IMU Evaluation Results

2.7 FriWalk Body Detector Module (FORTH)

The user perception modules on board the FriWalk must provide detailed information about the user in real time. Information such as the body posture, stability and gait of the user are valuable for both the clinical functions as well as the accident prevention and guidance capabilities of the device. As part of the user perception modules, FORTH has already developed and is currently improving a human body detector algorithm that can provide articulation information from marker-less visual observations obtained by a depth sensor. The method may operate with input provided by any of the on board depth sensors, mounted in the front and the back of the walker. Depending on the sensor used and on the distance of the user from the camera, the method can detect the user and estimate the 3D pose of either the full or the upper part of his/her body (up to 4 meters away). This capability enables the FriWalk to monitor the user's upper body while using the device and facilitates the implementation of clinical exercises and exergames.

Because of its high theoretical and practical interest, human motion capture based on vision has been the theme of numerous research efforts. The interested reader is referred to [15][16] where extended surveys are provided. More recently, Chen et al. [17] surveyed methods for human motion estimation based on depth cameras. Detecting the full articulation of the human body from visual observations is considered to be a difficult problem because of its high dimensionality and the variability of a body regarding appearance, dimensions, etc. A number of practical approaches simplify or even avoid these problems by using expensive, special hardware and/or by being invasive, e.g. by requiring that special visual or other markers are carefully placed on the human body to be tracked [14]. On the other hand marker-less human motion capture techniques may be classified into two broad classes, the bottom-up and the top-down ones. Bottom up methods [18][19][20][21][22] extract a set of features from the input images, and try to map them to the human pose space. This is achieved with a learning process that involves a typically large database of known poses that cover as much as possible the whole human poses search space. Top-down approaches [23][24][25][26][27][28] use a fully articulated model of the human body and try to estimate the joints angles that would make the appearance of this model fit best the visual input. The model is usually made of a base skeleton and an attached surface. A typical top-down method consists of generating hypotheses and comparing them to the input visual data. The comparison is performed based on an objective function that measures the discrepancy between a pose hypothesis and the actual observations. The main advantage of top-down methods is their extensibility. The employed model can be changed easily, and the whole search space can be explored without any form of training. The price to pay for this extensibility is the computational cost of the online process. Due to their generative nature, most of the computational work needs to be performed online. Two more shortcomings is the requirement for knowing the body model parameters of each individual and the requirement of providing an initial pose to be tracked.

According to the previous categorization, the method presented in this report is a hybrid with both top-down and bottom up elements. More specifically, hypotheses about 3D body parts are computed in a bottom-up approach but then refined and evaluated in a top down fashion. Moreover the presented method does not rely on tracking as most top-down approaches, and can perform body pose detection from single frames. FORTH's Body Detector has a number of important properties that are summarized as follows:

- performs accurate markerless detection of the human body in 3D
- achieves real time performance on a conventional computer
- requires simple inexpensive sensory apparatus (RGBD or depth camera)
- exhibits robustness in a number of challenging conditions (illumination changes, environment clutter, camera motion, etc)

- has a high tolerance with respect to variations in human body dimensions, clothing, etc. and
- can work in different modes such as upper body or full body configurations, thus it can perform even when the user is sitting or partially occluded.

The method employs a 3D articulated skeletal model of the human body. An illustration of this model is provided in Figure 23. This 3D model encapsulates information about the 3D positions of the human head, neck, shoulders, elbows, wrists, hips and legs as well as of the body center. For the purposes of the analysis, the body 3D model is hierarchically decomposed to the main body part B consisting of the head, shoulders, body center and hips and the limbs (arms and legs). A set of nine parameters (d1-d9) controls the sizes of body parts. The left/right symmetry of the human body is taken into account.

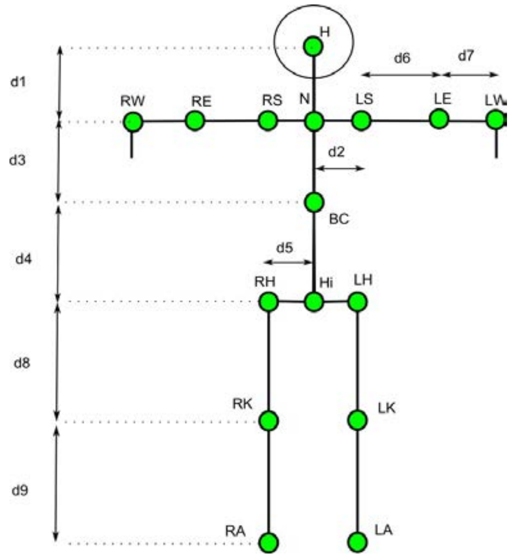


Figure 23: Body model and articulation.

The head is modeled as a spherical object centered in the head position. Arms are represented by two axis revolution volumes centered onto the shoulder-elbow and elbow-wrist 3D lines. The same applies to the body (neck, center, hip points). All model parameters related to sizes (length and radius of primitives) can assume values in predefined, broad ranges that cover most of the variability of human bodies, and are computed online. Several relations among these parameters are known because of anthropometric studies and taken into account in the estimation process. Thus, the evaluation of one parameter provides constraints on others.

The main steps of the method are shown in Figure 24. Starting from a raw depth map (top left) we threshold values further than 4 meters. This is required since with the current RGBD sensors the amount of noise beyond 4 m is significant. The second (preprocessing) step is to find 2D contours in the depth map, and perform skeletonization by applying erosion. The result of this process is a mask shown the top middle of the figure. Candidates for limb positions are then extracted from the preprocessed depth as well as the head position (top right). The candidate limbs and head position provide constrains for the body model. A final optimization step fits the body model to the input data. The best fitting candidate pose is shown super imposed in the RGB input on the bottom.

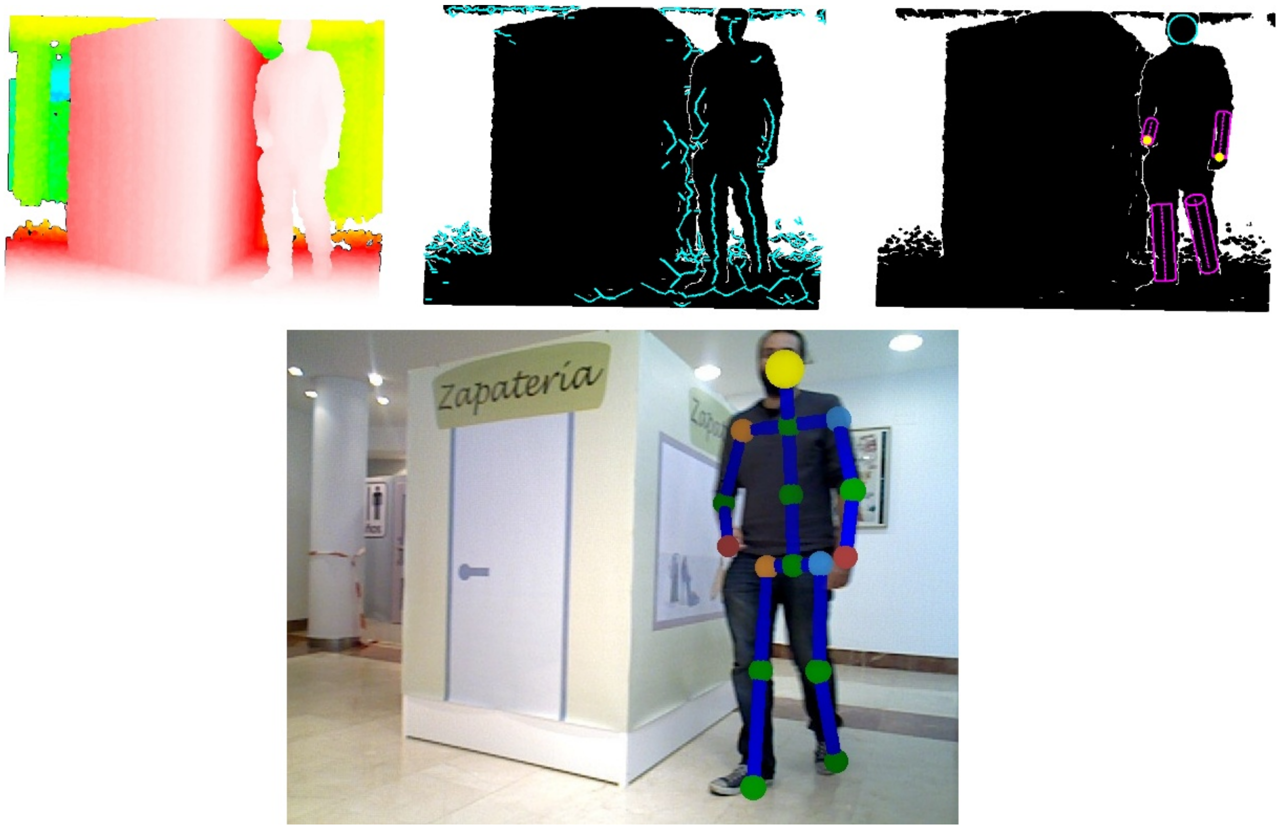


Figure 24: Skeleton detection steps. Starting from the raw depth map (top left) to the full body pose (bottom).

The method was tested in a number of scenarios where the user is either sitting or standing in front of the sensor. Figure 25 shows upper body detection with 3 different users with significantly different body shapes. From left to right, the input RGB image (left), the upper body model (middle) and the input depth map with the skeleton super imposed (right).

Typical results from the full body tracking are shown in Figure 26. The images shown are part of a dataset that was captured using the front mounted RGBD sensor of a DALi C-Walker. The position of the camera will be similar on the FriWalk platform. The body pose detection can provide results even if some body parts are not detected as long as the head and shoulders are visible. On the bottom left of each image the side view of the skeleton is shown.

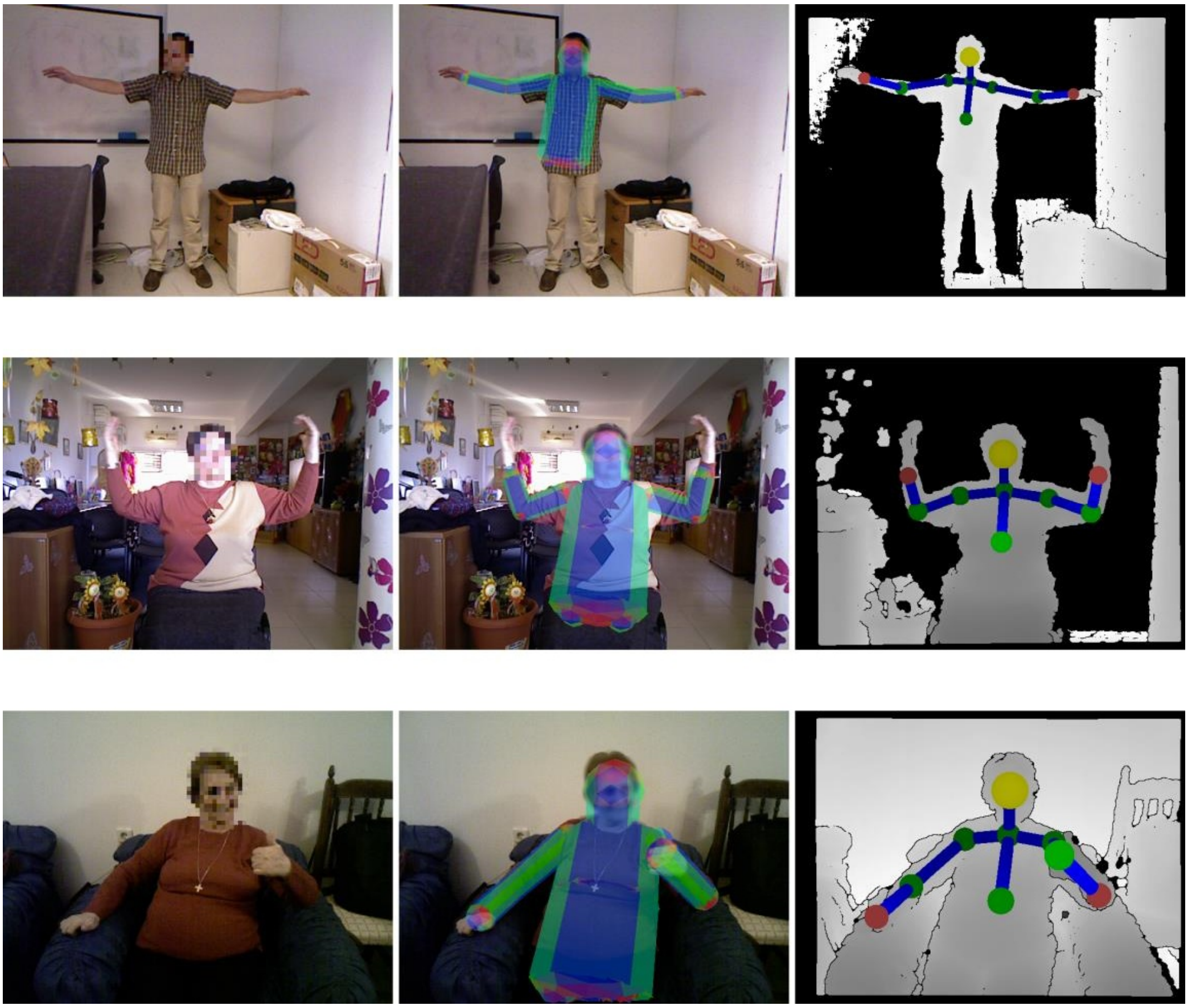


Figure 25: Upper body pose detection with different subjects.

Figure 26: Full body pose detection.

2.8 Conclusions on User-Centric Sensing

At the moment there is not a clear preference towards a specific COTS wearable device since all of them have some drawbacks, in particular with respect to continuous (real-time) data gathering and data synchronization, since the devices tend to report averaged values. On the other hand the market of wearables is continuously evolving and thus it is a good thing to have a wearable included in our system concept to allow for ongoing data acquisition even in the absence of the FriWalk as well as being able to measure aggregated values like the "*resting heart rate*".

This leaves room for tailored sensing on the FriWalk and requires research work to be carried out. Some approaches have shown to be rather promising like non-contact pulse measurement, gait analysis on the move and body pose detection. Other techniques like the non-contact respiration pattern measurement via radar failed to convince and does not seem applicable in the context of ACANTO.

Another system component that appears promising is the shoe insole sensor since it can reveal information about the stability of the gait that is not trivial or even impossible to access otherwise. However the deployment of such device in a later application scenario is more challenging than a sensor systems attached to or integrated into FriWalk. Hence it remains a topic for the future if such a device will be included in the final workflow (or selectively e.g. for instance in clinical applications) or it will mainly serve as ground truth reference like the Hexoskin smart garment.

Chapter 3
User state modelling (Siemens)

In this chapter, a first concept of the modelling architecture of the user state will be presented. Its hierarchical structures comprises of three levels of modelling (see Figure 27):

- The quantities of level 1, *"Vigilance"*, *"Activity index"* and *"Stress"*, are directly derived from the measurements discussed in chapter 2
- Level 2 is a semi-meta level: "*Medical indication*" and *"Emotional balance"* are calculated both from measurements and level1 quantities
- *"Overall Wellness"* is on level 3, a meta level, where all inputs are taken from previously calculated quantities

The different levels are presented in detail in the following subchapters.

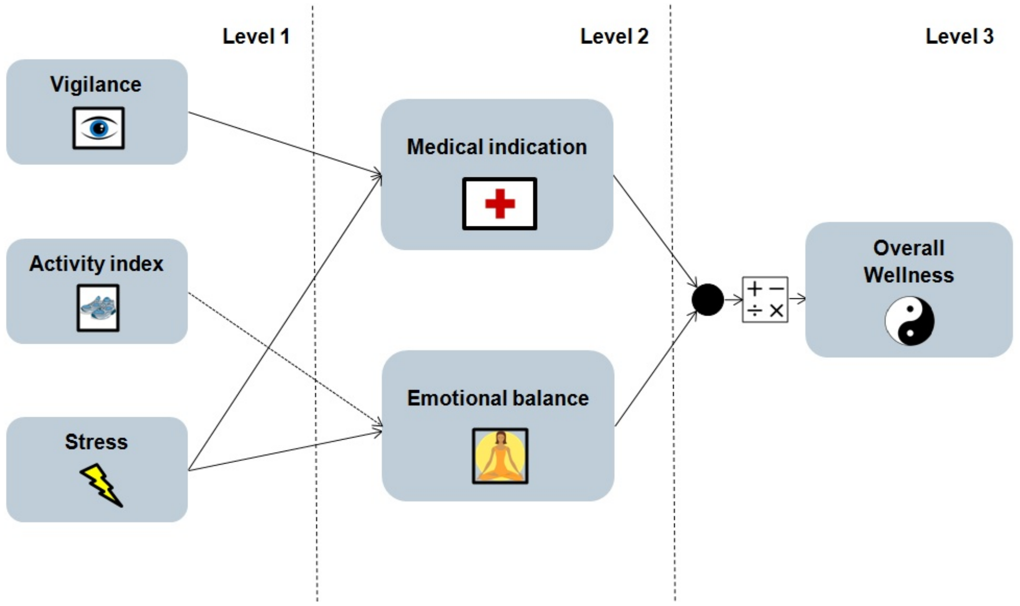


Figure 27: General model architecture of the user state model, which consists of three levels: the quantities *"Vigilance", "Activity index"* etc. are classified according to the origin of their input parameters. Level 1 means a direct calculation from measurements, Level 2 a calculation with both measured and previously calculated input, and Level 3 a calculation of previously calculated values only.

3.1 Level 1: "Vigilance" – "Activity Index" – "Stress"

3.1.1 Activity

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 paramount importance, we define the "*Activity index*" as a crucial field of the ACANTO user state modelling.

The input factors for the calculation of the "*Activity index*" are "*Time active*", "*Floors*", "*Steps/Distance*" (measured by the fitness tracker) and "*Heart rate*" (see section 2.4). In a first approximation, the "*Activity index*" will be measured in kilocalories burned by conducted activities. See Figure 28 for a graphical representation.

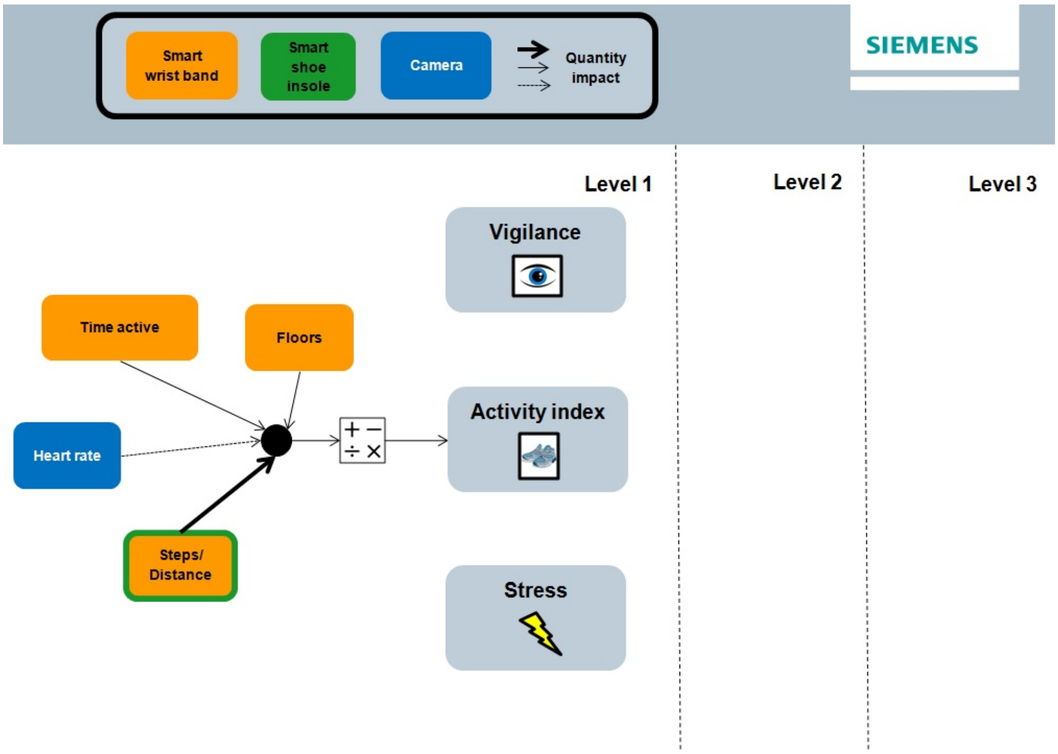


Figure 28: Sub-model architecture for the field "*Activity index*" of Level 1, which calculates from the quantities "*Heart rate*" (small impact), "*Time active*", "*Floors*" (medium impact), and "*Steps/Distance*" (high impact) of model Level 1.

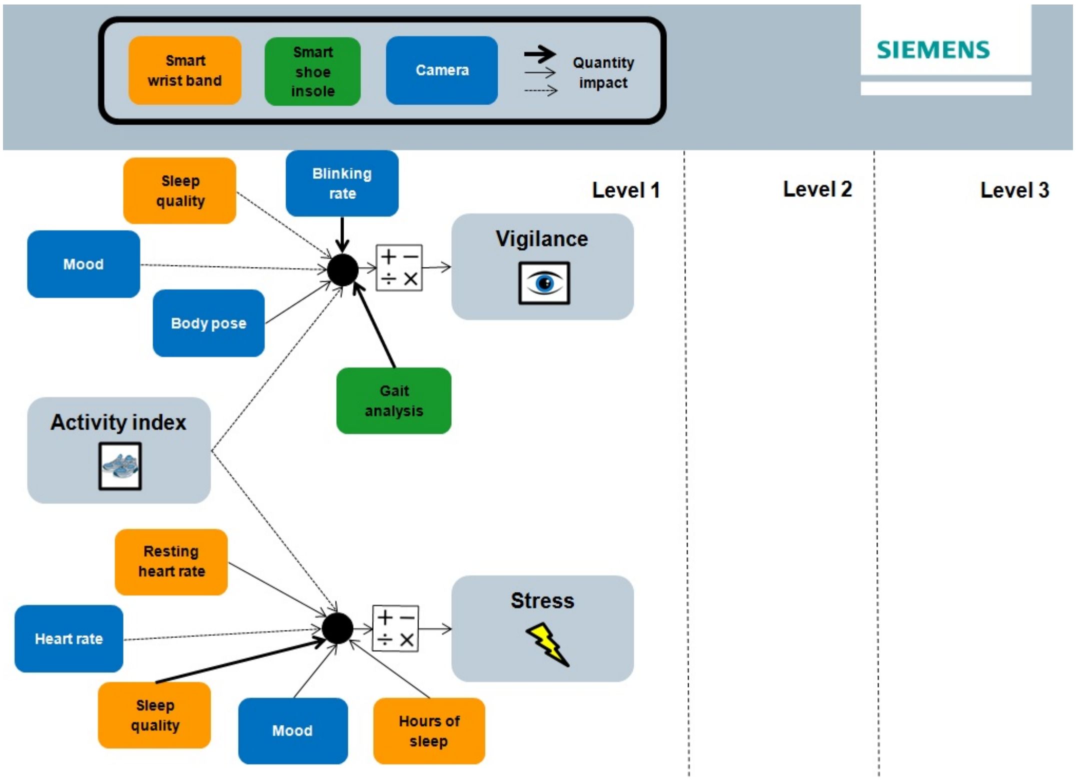


Figure 29: Sub-model architecture for the fields of Level 1, "*Vigilance*" and "*Stress*": the former calculates from the quantities "*Sleep quality*", "*Mood*", "*Activity index*" (small impact), "*Body pose*" (medium impact), "*Blinking rate*", "*Gait analysis*" (high impact); the latter calculates from "Heart *rate*", "*Activity index*" (small impact), "*Resting heart rate*", "*Mood*", "*Hours of sleep*" (medium impact), and "*Sleep quality*" (high impact).

3.1.2 Vigilance

"Vigilance" is another important field for the ACANTO project: An alert client perceives his environment better and suffers from fewer accidents. See Figure 29 for a sketch of input parameters.

"Blinking rate" is a proper means of measuring fatigue [32]. Within the project, this quantity is determined by the camera system. *"Sleep quality"* represents an important input quantity for "Vigilance" as it highly affects sleepiness and fatigue [33]. We define little change in *"Mood"* over time as indicator for decrease in *"Vigilance".* In a similar way, little change in *"Body pose"* might also be due to reduced *"Vigilance". "Gait analysis"* in general and the heel height during the single steps in particular might indicate psychic and psychical fatigue. We also see "*Activity index"* as a notable input factor for *"Vigilance"*.

3.1.3 Stress

Lazarus defines: *"Stress arises when individuals perceive that they cannot adequately cope with the demands being made on them or with threats to their well-being"* [34]. Stress is commonplace in today's life; exceeding arousal is linked to cardiovascular disease, cancer, arthritis, and major depression [35].

Stress and an elevated "*Resting heart rate*" (measured by the fitness tracker) are linked [36] and a high resting heart rate itself correlates with many kinds of mortality [37][39]. "*Resting heart rate*" is a long-term quantity, whose values must be observed over weeks. We define the current "*Heart rate*" as another input for stress. Unlike the "*Resting heart rate*", it can be seen as a short-term quantity indicating current stress situations.

Stress is associated with sleep disorders in two ways: stress provokes sleep disturbances, and disturbed sleep provokes stress and increases risk e.g. for cardiovascular disease [40]. Both "*Hours of sleep*" (about 6-9 hours per night [40]) and "*Sleep quality*" have to be considered [33]. In [38], sadness, lack of vigor, egotism, and social affection are linked to sleep deprivation. Another input factor for the field "*Stress*" is "*Mood*", which is investigated via the camera system. Excess of activity (measured as "*Activity index*") might lead to "*Stress*" as well.

3.2 Level 2: "Medical indication" – "Emotional balance"

3.2.1 Medical indication

"*Medical indication*" is defined as a short-term field and may trigger a warning to caregivers.

We define small values of "*Vigilance*", high levels of "*Stress*", or rapid alterations of "*Body pose*" and "*Heart rate*" – including superposition of these quantities – to possibly trigger a warning; see Figure 30 for a graphical representation.

3.2.2 Emotional balance

We define "*Emotional balance*" as a feeling of personal well-being without considering short-term "*Medical indication*". See Figure 30 for the corresponding sub-model architecture.

We proclaim that alertness ("*Vigilance*") correlates with "*Emotional balance*". The positive influence of activity ("*Activity index*") on "*Emotional balance*" is widely accepted [41]. Improvement of cognitive function in older adults [42] and reduction of depressive symptoms [43] serve as examples here. "*Stress*" also influences "*Emotional balance*" [45]. Poor "*Sleep quality*" was significantly correlated with increased physical health complaints and with increased feelings of tension, depression, anger, fatigue, and confusion" [33].

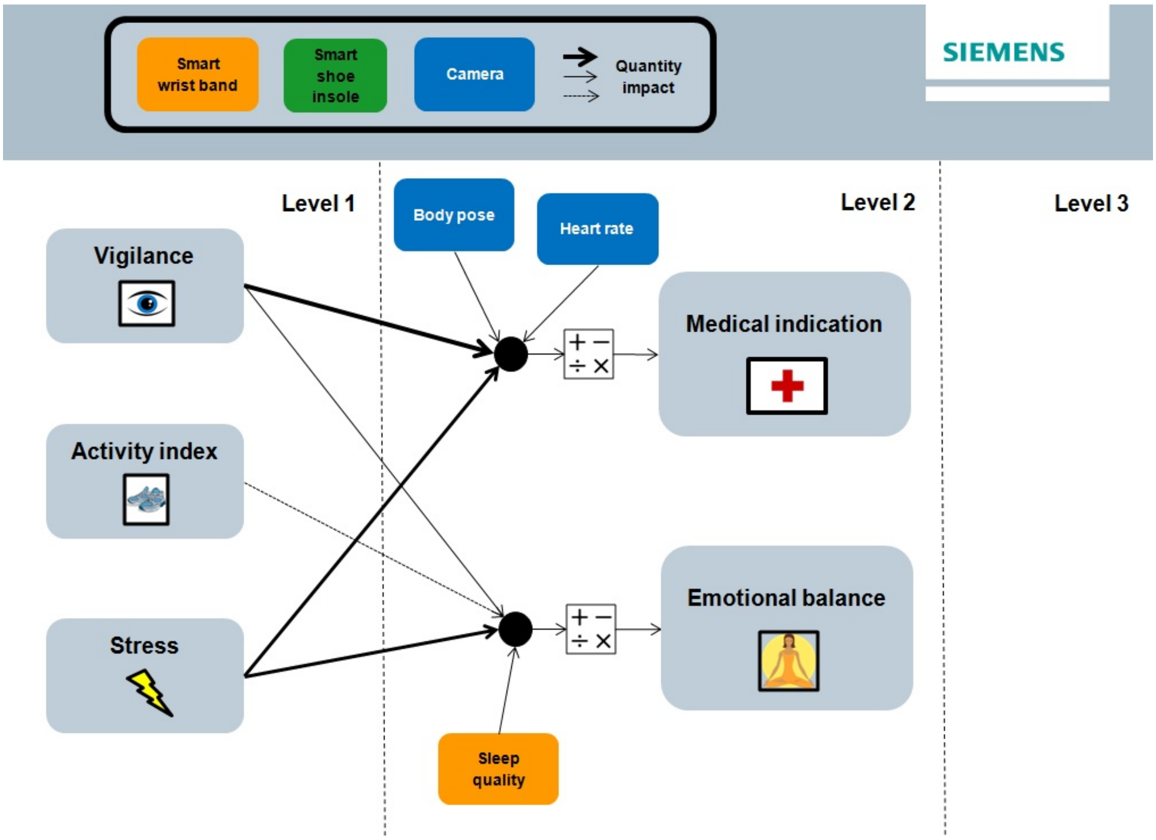


Figure 30: Sub-model architecture for the fields of Level 2, "*Medical indication*" and "*Emotional balance*": the former calculates from the quantities "*Body pose*", "*Heart rate*" (medium impact), "*Vigilance*", "*Stress*" (high impact); the latter calculates from "*Activity index*" (small impact), "*Vigilance*", "*Sleep quality*" (medium impact), and "*Stress*" (high impact).

3.3 Level 3: "Overall wellness factor"

As depicted in Figure 27, the "*Overall wellness factor*" comprises of the fields "*Medical indication*" and "*Emotional balance*" and represents a meta-quantity of the complete ACANTO user state model.

Chapter 4
Platform localization (UNITN)

4.1 Overview

Localization is essential to support navigation of FriWalks and robotic vehicles in general [46]. Usually, localization relies on both *relative* and *absolute* measurement techniques. In addition, sensor data fusion algorithms are commonly used to combine the benefits of both approaches [47]. The *relative* localization techniques measure the spatial and/or angular displacement of a robot or a vehicle with respect to a given initial position and/or attitude. Typical examples are the so-called dead reckoning methods based on odometers (for wheeled devices) or Inertial Measurement Units (IMU) [48]. The advantage of such techniques is their ability to self-measure the relative position and attitude of a robot at a low cost and at a high rate. Unfortunately, they also suffer from unbounded uncertainty accumulation and, particularly, systematic drifts. Moreover, the initial values of position and attitude are usually unknown.

On the other hand, the *absolute* localization techniques estimate the position and/or the attitude of a target within a global reference frame. The main advantage of such techniques is that the positioning and attitude uncertainties remain bounded whenever measurement data are available. However, multiple systems in the environment or a monitoring infrastructure are needed to locate the target. Therefore, various problems of cost, detection range, communication bandwidth, robustness and scalability may arise. In the case of outdoor vehicles, absolute localization is generally provided by the Global Positioning System (GPS). In addition, the absolute robot heading can be measured through a magnetometer used as compass. However, both GPS receivers and magnetometers can be hardly used indoors.

For all the reasons above, different measurement techniques as well as various ad-hoc data fusion strategies have been developed to estimate the position of multiple FriWalks in indoor environments. In particular, the overall localization strategy is twofold and consists of various functional blocks, as shown in Figure 31. In fact, localization relies on both local position values estimated by each FriWalk, and data from (or relative to) other robots (also referred to as agents in the following), whenever they are available. Locally, each FriWalk is equipped with a set of propriosensors that allows the robot to self-estimate its own position autonomously. In order to keep position accuracy steadily within ±1 m, two position estimation algorithms have been developed and compared, i.e. an Extended Kalman Filter (EKF) and an Extended H∞ Filter (EHF). Both algorithms stem from the general idea of merging vision-based measures and dead reckoning techniques. Indeed, if the landmarks are easy to detect and if their density in the environment is reasonably low, the data fusion of vision-based measurements and dead reckoning is a viable solution to achieve accurate, scalable and trustworthy localization [49], [50]. In the framework of ACANTO a coarse-grained grid of landmarks consisting of Quick Response (QR) codes stuck on the floor is actually used. The main advantage of using QR codes as landmarks is that both absolute position and heading can be measured in one shot and with the same delay, thus assuring good temporal data alignment.

The idea of using and comparing the performances of EKF and EHF is due to the fact that the vision-based position and heading measurement errors do not meet the typical assumptions of Kalman filters, since such errors are not white and normally distributed. As a result, the EKF is no longer optimal and its accuracy becomes hardly predictable. On the contrary, the EHF does not requires any a-priori knowledge of either noise distribution or shape of the power spectral density. Also, an EHF is purposely designed to minimize the worst-case estimation error. The difference between both approaches and a performance comparison based on experimental data is reported in Section 4.2.

It is important to emphasize that, when multiple FriWalks are present in the same environment, localization requirements become stricter than in the case of a single agent (i.e. within ±50 cm), since the position errors have to be small enough to prevent collisions between different walkers.

In order to achieve such a challenging goal, a collaborative localization approach is envisioned. The basic idea is to exploit the intrinsic *social* nature of the FriWalks. Byy sharing the information of their position in a common reference frame and by measuring the distance and/or the relative orientation between pairs of nearby agents, the position of each user can be refined. To this purpose in Section 4.3, it is explained how the local estimators can be enhanced to include the mutual information between pairs of agents. Also, two different inter-robot relative measurement schemes will be analyzed and compared through simulations, i.e. using either a large number of highly probable and low-accuracy wireless distance values or a small number of vision-based (and more accurate) pose measures.

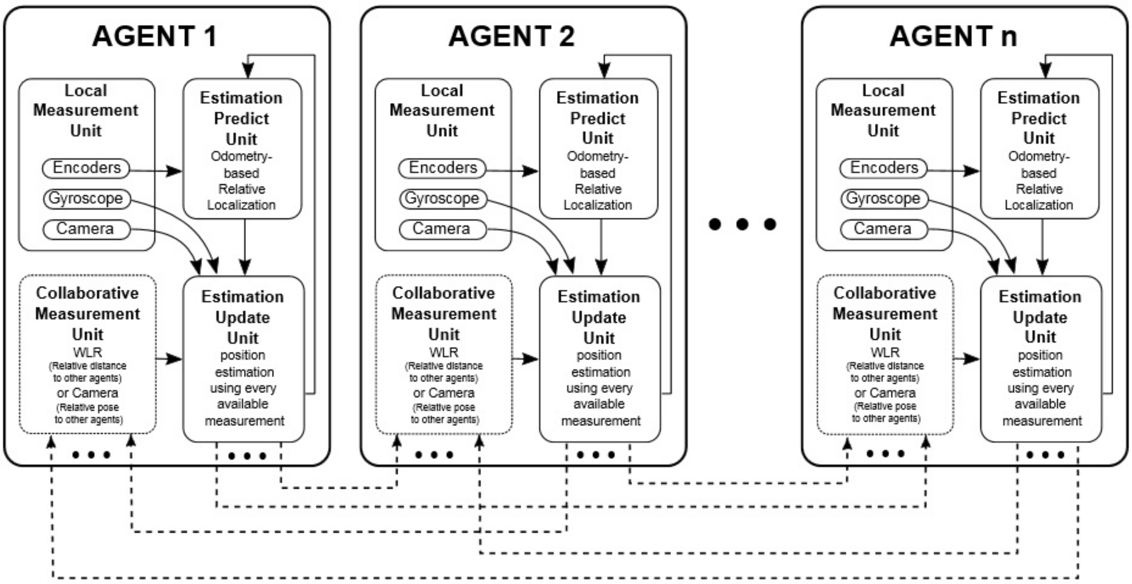


Figure 31: Block diagram of the overall localization strategy developed for the ACANTO FriWalks. The collaborative measurement unit is active only when multiple agents are present in the same environment.

4.2 Self-localization of a single agent

4.2.1 System and measurement models

A qualitative overview of the self-localization mechanism of a single agent is shown in Figure 32. A FriWalk equipped with two encoders, a gyroscope and a front monocular camera must be able to estimate its own position within a reference frame <*W*>={X_w, Y_w, Z_w}. The robot's generalized coordinates at time kT_s (T_s being the sampling period) compose the state vector $\boldsymbol{p}_k = [x_k, y_k, \theta_k]^T$, where ($x_k$, y_k) are the planar coordinates of the mid-point of the back wheels axle (in the following simply referred to as *reference point* for brevity) and θ_k is the heading of the robot with respect to X_w. The FriWalk dynamic can be modeled as a unicycle-like vehicle, i.e.

$$\begin{aligned} \boldsymbol{p}_{k+1} &= \boldsymbol{p}_k + G_k(\boldsymbol{u}_k + \boldsymbol{\varepsilon}_k) \\ \boldsymbol{z}_k &= \boldsymbol{o}(\boldsymbol{p}_k) + \boldsymbol{\eta}_k \end{aligned} \tag{1}$$

where $\boldsymbol{u}_k = [\delta s_k, \delta\theta_k]^T$ is the input vector including the linear and angular displacements of the vehicle between (k-1)T_s and kT_s, $\boldsymbol{\varepsilon}_k$ is the vector including the respective zero-mean noise terms (assuming that possible systematic offsets are reasonably estimated and compensated),

$$G_k = \begin{bmatrix} \cos\theta_k & 0 \\ \sin\theta_k & 0 \\ 0 & 1 \end{bmatrix} \qquad (2)$$

$\boldsymbol{o}(\boldsymbol{p}_k)$ denotes a generic nonlinear output function of the state and $\boldsymbol{\eta}_k$ is the vector of the zero-mean uncertainty contributions, when the output quantities are measured.
If the position of a robot that started from a known location is estimated by integrating the wheels displacements measured by the encoders, the accumulation of both imperfectly compensated offsets and random noise unavoidably leads to large position and heading errors after a while. In order to mitigate this problem, two measurement systems are included in the FriWalk, i.e.

- A vision system consisting of a monocular front camera;
- A gyroscope-based platform.

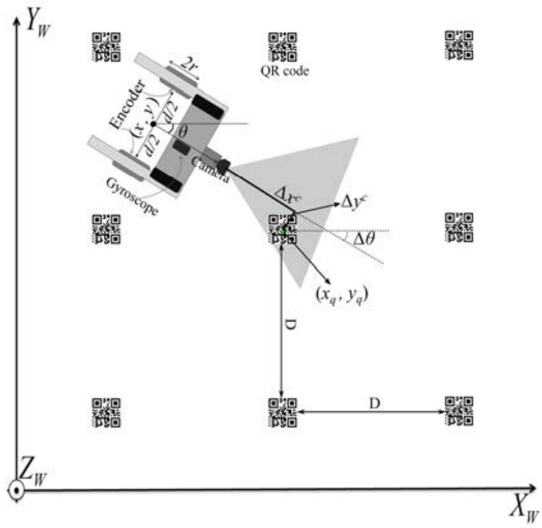


Figure 32 – Overview of the localization mechanism of a single agent within the fixed reference frame <W> = {Xw. Yw. Zw}.

The role of the vision system is to detect one of the QR codes stuck on the floor. To this purpose, the field of view of the camera has to be slightly oriented towards ground in order to ensure a front reading range of at least 1 m. One of the main advantages of using QR codes is that they can be robustly identified in the environment. Also, various open-source libraries and algorithms are available to read such codes (e.g. Zbar) and to measure both the distance of a detected landmark from the camera and its orientation angle with respect to the optical axis [52]. In the following, for the sake of simplicity, but without loss of generality, it will be assumed that the distance D between pairs of adjacent QR codes is constant, as shown in Figure 32. Each QR code represents a number which corresponds to a triple of values, i.e. the planar coordinates (x_q, y_q) of the point in the center of the landmark, and its orientation angle θ_q with respect to X_w (by default, θ_q=0).

As soon as the q-th QR code is recognized, the values of x_q, y_q and θ_q are extracted from a list of QR code numbers stored in the memory of the platform running the localization algorithm. This approach is very flexible, since the same QR codes can be used in different environments by simply remapping the numbers in the list. Also, in this way just low-resolution (e.g. 8-bit) numeric-only codes can be used. As a result, the probability of correct code detection is higher even at larger distances. With reference to Figure 32, let Δx^c and Δy^c be the distances between the camera and the detected QR code in the camera frame, namely along its optical axis and the orthogonal direction parallel to ground, respectively. Also, let $\Delta\theta^c$ be the angle difference between the camera optical axis and θ_q. The values of such quantities can be measured using standard image processing algorithms, e.g. based on homography [53]. If the points of the landmark to detect are coplanar (like in the case of the QR codes) and if the geometrical dimensions of such codes are known a-priori, the homography-based techniques ensure a robust estimation of Δx^c, Δy^c and $\Delta\theta^c$, regardless of the actual position and orientation of the camera. If $\Delta\hat{x}^c = \Delta x^c + \eta_x$, $\Delta\hat{y}^c = \Delta y^c + \eta_y$ and $\Delta\hat{\theta}^c = \Delta\theta^c + \eta_\theta$ denote the measured quantities, the distributions of uncertainty terms η_x, η_y and η_θ depend on the combination of a multitude of phenomena such as possible camera calibration errors, algorithm-specific uncertainty contributions introduced by the image processing stage, intrinsic pose errors of the landmarks and, last but not least, random latencies due to the fact that images are processed while the robot is moving. The relative position and angle measurements obtained from the vision system are inherently event-based. In fact, only when one of the QR codes is in the field of view of the camera, the related information is available. So, not only is the detection rate random, but it is also a function of the density of the landmarks in the environment. Given that ensuring high-rate and good heading estimates can greatly improve localization accuracy in the long term [54], an additional gyroscope-based platform can be used. As known, the angular displacements around a given rotation axis can be obtained by simply integrating the values measured by a gyroscope (around the same axis) over time. This approach generally also suffers from uncertainty accumulation. However, the vision-based angle measurements can be used in an additional EKF to adjust the yaw angular displacements resulting from the integration of gyroscope data, thus keeping the total heading uncertainty bounded. For the aims of ACANTO, potentially the rotation axis of the gyroscope should be orthogonal to ground, i.e. parallel to Z_w. If this is not the case, the angular velocity components measured by a triaxial gyroscope in its own reference frame have to be properly combined so as to compute the components in frame <W>. This requires estimating the attitude of the gyroscopic platform within <W>. Such attitude estimates can be obtained by means of an EKF in which the direction of the gravity vector is measured by a triaxial accelerometer while the direction of magnetic north is determined by a triaxial magnetometer. The details of this approach are explained in [55]. Therefore, in the following, it will be simply assumed that the angular velocity component along Z_w is measured as described in [55]. It is important to emphasize instead that the potential poor accuracy of magnetometer measurements in indoor environments is not an issue in the case at hand, since it does not affect the angular velocity component along Z_w.

4.2.2 Local position estimation algorithms

The heterogeneity of the chosen sensors as well as the adopted nonlinear system model require an efficient algorithm to coherently fuse the available data and to achieve the best possible estimates of the elements of $\boldsymbol{p}_k$ in (1). As explained in Section 0, the successful application of Kalman filtering techniques depends on the fulfilment of prerequisites on process and measurement noises, regardless of the linearity of the underlying system. More precisely, such noises should be white and normally distributed with zero mean and known covariance. Such hypotheses are hardly verified in real scenarios, particularly for the case at hand. In addition, the hypothesis of white noise is evidently not valid for the heading measurements resulting from the gyroscope-based EKF shortly described in the previous section, since the estimated values are highly correlated. In this kind of situations, H_∞ filters can improve robustness to unmodelled noise and dynamic phenomena. In order to highlight the differences between the EKF and the EHF based on the same system model (1), both solutions are described and compared in the following.
As far as the EKF is concerned, it can be easily shown that the equations of the prediction step are

$$\begin{aligned} \hat{\boldsymbol{p}}_{k+1}^{+} &= \hat{\boldsymbol{p}}_k + G_k\, \boldsymbol{u}_k \\ P_{k+1}^{+} &= F_k P_k F_k^T + G_k Q_k G_k^T \end{aligned} \tag{3}$$

where $\hat{\boldsymbol{p}}_k$ and $\hat{\boldsymbol{p}}_{k+1}^{+}$ denote the estimated and predicted state, respectively, P_k and P_{k+1}^{+} are the corresponding covariance matrices, F_k is the Jacobian of the system model (1) with respect to $\boldsymbol{p}$ and computed at $\hat{\boldsymbol{p}}_k$, G_k is defined as in (2), and, finally, Q_k is the covariance matrix of the process noise $\boldsymbol{\varepsilon}_k$ mainly due to encoders. If we refer to O_{k+1} as the Jacobian of the output function $\boldsymbol{o}(\cdot)$ in (1) with respect to $\boldsymbol{p}$ and computed at $\hat{\boldsymbol{p}}^{+}_{k+1}$, then the Kalman gain of the EKF is

$$K_{k+1} = P^{+}_{k+1} O^{T}_{k+1} \left(O_{k+1} P^{+}_{k+1} O^{T}_{k+1} + R_{k+1} \right)^{-1} \tag{4}$$

where R_{k+1} is the covariance matrix of $\boldsymbol{\eta}_{k+1}$. Observe that the size of both O_{k+1} and R_{k+1} depends on whether a QR code is read or not. In the former case, O_{k+1} is a 3x5 matrix, while the size of R_{k+1} is 3x3. In the latter case, O_{k+1} is a 1x5 row vector and R_{k+1} is a scalar. Of course, the size of the observation vector $\boldsymbol{z}_{k+1}$ changes as a function of the available measures as well. Nonetheless, in both cases the update step equations of the EKF are

$$\begin{aligned} \hat{\boldsymbol{p}}_{k+1} &= \hat{\boldsymbol{p}}^{+}_{k+1} + K_{k+1}\left(\boldsymbol{z}_{k+1} - \boldsymbol{o}\left(\hat{\boldsymbol{p}}^{+}_{k+1}\right)\right) \\ P_{k+1} &= \left(I_3 - K_{k+1} O_{k+1}\right) P^{+}_{k+1} \end{aligned} . \tag{5}$$

As far as the EHF is concerned, this results from the minimization of the following cost function [51]:

$$J = \frac{\sum_{j=0}^{k} \left\| L_j\, \boldsymbol{p}_j - L_j\, \hat{\boldsymbol{p}}_j \right\|^2}{\left\| \boldsymbol{p}_0 - \hat{\boldsymbol{p}}_0 \right\|^2_{P_0} + \sum_{j=0}^{k} \left\| \boldsymbol{\varepsilon}_j \right\|^2_{\left(G_j Q_j G_j^T\right)^{-1}} + \sum_{j=0}^{k} \left\| \boldsymbol{\eta}_j \right\|^2_{R_j^{-1}}} \tag{6}$$

where P_0, Q_j, R_j, and L_j are positive definite weighting matrices which should be chosen appropriately. In general, the exact solution of the optimization problem above is not tractable. Therefore, in practice it is preferable to compute the minimum of (6) such that $\boldsymbol{sup}_k\, J < \gamma^2$, with γ being an arbitrary given boundary. In fact, the optimal solution to this constrained optimization problem is easier to find. By following the steps of the algorithm proposed in [56], it can be indeed shown that the prediction equations are the same as (3), whereas the equations of the update step become

$$\begin{aligned} \hat{\boldsymbol{p}}_{k+1} &= \hat{\boldsymbol{p}}^{+}_{k+1} + K_{k+1}\left(\boldsymbol{z}_{k+1} - \boldsymbol{o}\left(\hat{\boldsymbol{p}}^{+}_{k+1}\right)\right) \\ P_{k+1} &= \left(I_3 - P^{+}_{k+1} \begin{bmatrix} O^{T}_{k+1} & L^{T}_{k+1} \end{bmatrix} B^{-1}_{k+1} \begin{bmatrix} O_{k+1} \\ L_{k+1} \end{bmatrix} \right) P^{+}_{k+1} \end{aligned} . \tag{7}$$

where the H_∞ gain is computed as in (5) and

$$B_{k+1} = \begin{bmatrix} R_{k+1} & 0 \\ 0 & -\gamma^2 I \end{bmatrix} + \begin{bmatrix} O_{k+1} \\ L_{k+1} \end{bmatrix} P^{+}_{k+1} \begin{bmatrix} O^{T}_{k+1} & L^{T}_{k+1} \end{bmatrix}. \tag{8}$$

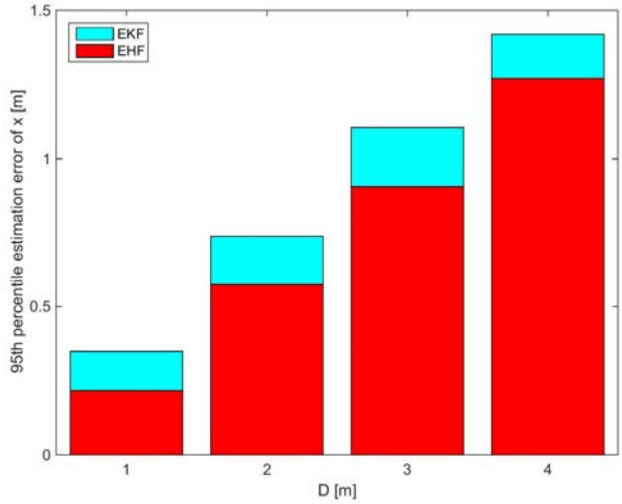
This way of formalizing the EHF makes its representation very similar to an EKF, although its computational cost is higher. However, Q_k and R_k are no longer required to be the covariance matrices of the process and measurement noises, respectively. In fact, they can comprise arbitrary user-defined weights, with the only constraint that both matrices have to be positive definite. If some a-priori knowledge is available, Q_k and R_k can be used to leverage the importance of different kinds of noise. Nonetheless, it is worth emphasizing that if Q_k and R_k coincide with the covariance matrices of process and measurement noises, respectively, and if L_k is equal to the identity matrix, then the EHF tends to become an EKF when $\gamma \rightarrow \infty$.

4.2.3 Implementation and results

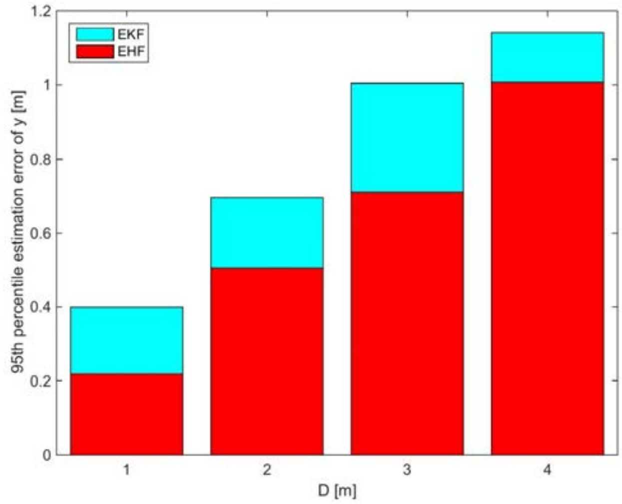
The behavior of both kinds of filters has been evaluated and compared using several records of data collected on the field using a preliminary prototype of the robotic walker. The wheels of the prototype are equipped with two incremental optical encoders CUI Inc. AMT10X with a best nominal accuracy of ±15 arcmin at the highest possible resolution (i.e. 2048 PPR). The robot's wheels radius is 10 cm and the axle length is 59 cm long. The adopted gyroscope is a triaxial Inversense IMU-3000 with a resolution of 16 bits and a declared rate noise spectral density of $1.74 \cdot 10^{4}$ rad/s/√Hz. The attitude of the gyroscope within frame <*W*> is obtained as described in [55], by using a 14-bit triaxial accelerometer Bosch BMA180 to estimate the direction of Z_w, and a 12-bit triaxial magnetometer Honeywell HMC5883 to estimate the direction of X_w. Both encoders and gyroscope data are collected via CAN bus at 250 Hz (i.e. with period T_s = 4 ms). The front camera is a simple 640x480 PSeye RGB webcam connected to the processing platform through a USB port. The camera is placed at a height of about 80 cm and 60 cm ahead of the reference point (namely the wheel's axle midpoint) along the longitudinal axis of the robot. The processing platform is an Intel Barebone mini desktop (of size 11.7 cm x 11.2 cm x 3.9 cm) provided with 8 GB of DDR3 RAM, a 2.80-GHz Intel Core i5-3427U processor, a 120-GB solid state drive and Linux embedded. The processing platform is powered by a rechargeable 118-Wh lithium-ion accumulator. Landmark recognition is implemented in C++ using the primitives of the OpenCV library. The open-source Zbar library is used instead for QR code detection. Both the EKF and the EHF filter are implemented in C.

Some experiments were conducted in a large room of about 300 m^2 in the basement of the Department of Industrial Engineering of the University of Trento. The room was instrumented with QR markers put on the floor at distances D=1, 2, 3 or 4 m from one another. About 40 experiments of different duration were conducted with the FriWalk prototype pushed at various speeds and along different routes, both throughout the empty room and with some obstacles on the way to emulate realistic scenarios. In order to evaluate the accuracy of the position tracking technique, a laser scanner SICK S300 Expert was placed in the origin of reference frame <*W*> (i.e. in one corner of the room) to measure the coordinates of the user along each route in real-time.

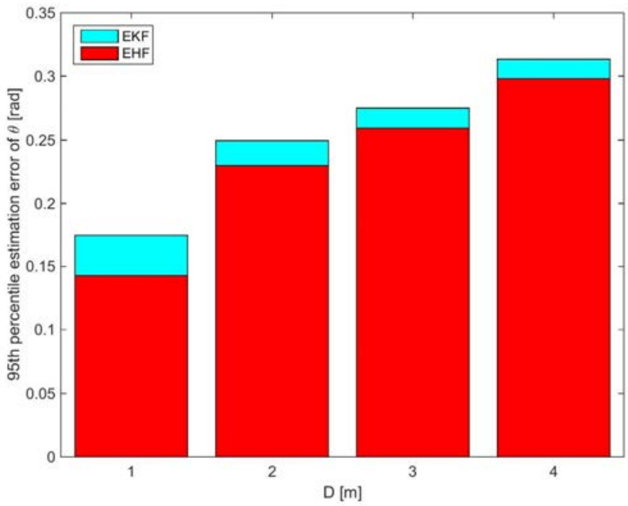
Figure 35 shows the 95th percentiles of the estimation errors of state variables x (a), y (b) and θ (c) obtained using the EKF and the EHF, respectively, with QR code grids of different granularity. Of course, as the distance D between landmarks grows, the estimation errors tend to increase considerably, as the probability of detecting a landmark along the way becomes significantly smaller. Observe that in all conditions, the accuracy achieved with the EHF is higher than with the EKF. In particular, in order to have a positioning error within ±1 m with about 95% probability, the distance D between pairs of adjacent landmarks must be at least 2 m. In this case, the accuracy improvement obtained using the EHF is between 20% and 25% for state variables x and y, while it is about 10% for variable θ.



(a)

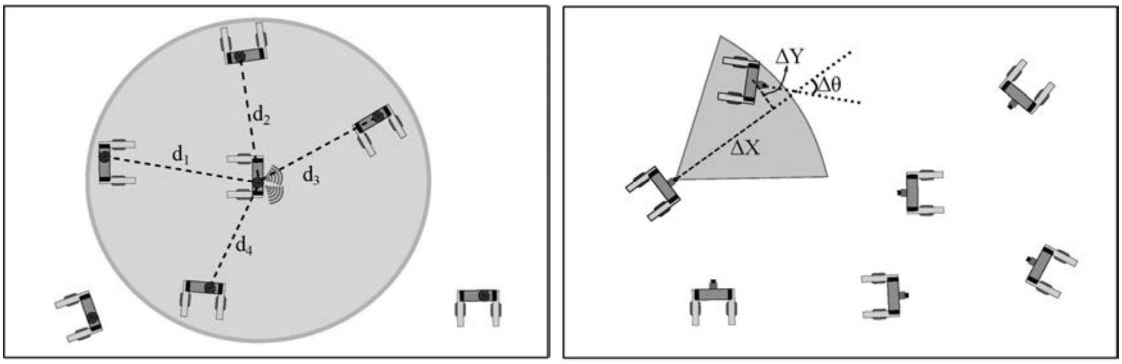


(b)



(c)

Figure 33: 95th percentiles of the estimation errors of state variables *x* (a), *y* (b) and *θ* (c) obtained using the EKF and the EHF, respectively, in about 40 experiments, for different values of the distance *D* between pairs of QR code landmarks.



(a) (b)

Figure 34: Qualitative overview of measurement techniques used for collaborative localization of multiple FriWalks, i.e. omni-directional wireless ranging systems (distance measurement only) (a), Kinect cameras measuring the relative position and orientation of two devices (b).

4.3 Collaborative localization

As explained in Section 0, the accuracy of the self-localization performed by a single agent, even if it plays an essential role for most of the functions of the FriWalk, could be not good enough to support the simultaneous and coordinated maneuvering of different robots in the same room. In order to tackle this problem, the possibility of using collaborative localization has been investigated both analytically and through some preliminary simulations. Till now, all efforts have been focused on the idea of injecting mutual inter-robot measurements in the update step of each EKF in order to turn them into a distributed and interlaced EKF (IEKF) [57]. A similar approach could be potentially applied also to the EHF. However, its complex theoretical formulation and possible stability issues require a longer dedicated study, which will be performed in future if needed.

4.3.1 Problem formalization

For *collaborative* or *synergic* localization of a team of agents we mean the ability to refine the position and the heading estimated by each agent in a common reference frame by using both local positioning data and relative distance and/or orientation measures between pairs of devices. The main assumptions underlying a proper formalization of this problem in the specific context of the ACANTO project, are summarized below.

1. The N agents can move freely in a large room. In other words, the dynamic of each agent does not depend on any other agent, since each user may act independently. The only constraint to FriWalk motion is collision avoidance.
2. The state of each agent i (with i=1,…,N) at time kT_s is represented by vector $p_k^{(i)}=[x_k^{(i)}, y_k^{(i)}, \theta_k^{(i)}]^T$, where $x_k^{(i)}$, $y_k^{(i)}$ and $\theta_k^{(i)}$ are defined like in (1).
3. Each FriWalk is able to estimate its own state autonomously as explained in Section 4.2.
4. Besides the sensors used by each robot for its own local state estimation, every FriWalk is supposed to be equipped with two alternative types of exteroceptive sensors to be used explicitly for collaborative localization, i.e. an omni-directional wireless ranging system (*case A*) or a front stereo vision system (e.g. a Kinect) (*case B*). Both cases are qualitatively shown in Figure 36 (a) and (b), respectively. In *case A* the ranging system is used to measure just the distance between the FriWalk and any other agent located within an (approximately) circular range. On the contrary, in *case B* the Kinect is employed to recognize and to measure the relative position and orientation between the FriWalk and any other agents located within both a given horizontal angle of view and a known min-max range.

5. All agents are equipped with a radio transceiver ensuring complete connectivity between any pairs of robots as well as high-rate and low-latency communication. Long Term Evolution (LTE) wireless modules can indeed meet such general requirements [58].

By extending model (1), the overall state transition of all agents in the chosen reference frame can be described by the following non-linear discrete-time system [59], i.e.

$$\boldsymbol{s}_{k+1} = \begin{bmatrix} \boldsymbol{p}_{k+1}^{(1)} \\ \vdots \\ \boldsymbol{p}_{k+1}^{(N)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{p}_k^{(1)} + G_k^{(1)}\left(\boldsymbol{u}_k^{(1)} + \boldsymbol{\varepsilon}_k^{(1)}\right) & \cdots & 0 \\ \vdots & \cdots & \vdots \\ 0 & \cdots & \boldsymbol{p}_k^{(1)} + G_k^{(1)}\left(\boldsymbol{u}_k^{(1)} + \boldsymbol{\varepsilon}_k^{(1)}\right) \end{bmatrix} \qquad (9)$$

where each term of the block diagonal matrix in (9) describe the dynamic of one of the N FriWalks in the room. If we assume to have inter-robot mutual measurements, the observation model associated to system (1) includes two types of output functions, i.e.

- the geometrical relationship between the position/orientation of each agent and those of one of the detected visual landmark in a common reference frame;
- the geometrical relationship between the pose of each FriWalk and those of the other N-1 agents in the room.

As a result, the overall observation equation at time kT_s becomes

$$\boldsymbol{z}_k = \begin{bmatrix} \boldsymbol{z}_k^{(1)} \\ \vdots \\ \boldsymbol{z}_k^{(N)} \end{bmatrix} = \begin{bmatrix} \tilde{\boldsymbol{h}}_k^{(1)}(\boldsymbol{s}_k) \\ \vdots \\ \tilde{\boldsymbol{h}}_k^{(N)}(\boldsymbol{s}_k) \end{bmatrix} + \begin{bmatrix} \boldsymbol{\eta}_k^{(1)} \\ \vdots \\ \boldsymbol{\eta}_k^{(N)} \end{bmatrix} \qquad (10)$$

where $\boldsymbol{z}_k^{(i)} = \left[\boldsymbol{z}_k^{(i,1)}, \ldots, \boldsymbol{z}_k^{(i,N)}\right]^T$ includes all possible observations from agent i, $\boldsymbol{\eta}_k^{(i)} = \left[\boldsymbol{\eta}_k^{(i,1)}, \ldots, \boldsymbol{\eta}_k^{(i,N)}\right]^T$ is the vector comprising the respective measurement uncertainty contributions and, finally,

$$\tilde{\boldsymbol{h}}^{(i)}(\boldsymbol{s}_k) = \begin{bmatrix} \boldsymbol{h}\left(\boldsymbol{p}_k^{(i)}, \boldsymbol{p}_k^{(1)}\right) \\ \vdots \\ \boldsymbol{h}\left(\boldsymbol{p}_k^{(i)}, \boldsymbol{p}_k^{(1-1)}\right) \\ \boldsymbol{o}\left(\boldsymbol{p}_k^{(i)}\right) \\ \boldsymbol{h}\left(\boldsymbol{p}_k^{(i)}, \boldsymbol{p}_k^{(1+1)}\right) \\ \vdots \\ \left(\boldsymbol{p}_k^{(i)}, \boldsymbol{p}_k^{(N)}\right) \end{bmatrix} \qquad i=1,\ldots,N, \qquad (11)$$

is the vector including all observations performed by agent i. Observe that the i-th function of the vector, referred to as $\boldsymbol{o}(\cdot)$, is the same as in (1) and it is different from the other elements, as it depends on the geometrical relationship between the position/orientation of each agent and those of the detected QR code. On the contrary, each function $\boldsymbol{h}(\cdot,\cdot)$ consists of M equations and depends on how the state variables of agents $j = 1,\ldots,N$, for $j \neq i$ are actually observed by the i-th robot. In the following, it will be assumed that while $\boldsymbol{o}(\cdot)$ is the same in all conditions (see assumption 3), the equations of $\boldsymbol{h}(\cdot,\cdot)$ differ in *case A* and *case B*, respectively, in accordance with assumption 4.

4.3.2 Collaborative position estimation algorithm

The nonlinear model based on expressions (9)-(10) can be used to implement an estimation algorithm similar to the IEKF presented in [59]. Since the state evolution of each FriWalk does not depend on the state of the other robots, the prediction step equations of the IEKF are straightforward, as they basically coincide with those of N independent EKFs, i.e.

$$\begin{aligned} \hat{\boldsymbol{p}}_{k+1}^{+(i)} &= \hat{\boldsymbol{p}}_{k}^{(i)} + G_k^{(i)} \boldsymbol{u}_k^{(i)} \\ P_{k+1}^{+(i)} &= F_k^{(i)} P_k^{(i)} F_k^{(i)T} + G_k^{(i)} Q_k^{(i)} G_k^{(i)T} \end{aligned} \qquad i=1,\dots,N, \qquad (12)$$

where the meaning of all symbols is the same as in (3), but referred to the i-th agent. Observe that (12) depends just on local quantities. Therefore, the prediction step equations can be computed locally, i.e. on board of each FriWalk, thus ensuring a fully distributed implementation.

As far as the update step is concerned, the corresponding equations are different from those of a standard EKF for two reasons. First of all, due to the definition of the measurement model (10), the updated state estimate of the i-th FriWalk and its covariance matrix depend not only on the respective predicted values and on the measurement data, but also on the predicted state and the predicted covariance of the other agents. Secondly, generally a FriWalk is not able to observe all the other agents simultaneously and can occasionally miss landmarks as well. This means that all observations are inherently intermittent, as they depend on the reading range of the chosen measurement systems and on the distance between each FriWalk and both the other agents and one of the landmarks. In particular, if the distance between pairs of adjacent landmarks is larger than the reading range of the camera, then at most one landmark can be used to update the state of each agent. As a result of all issues above, in the case considered the update step of the IEKF is inherently stochastic.

The possibility to implement a local state estimator depending also on the states of other agents is investigated in [59], and led indeed to the idea of using an IEKF. In essence, this means that in the computation of the innovation term associated with a generic pair of agents (i,j), the predicted state of j can be regarded as an additional measure. Therefore, both $\hat{\boldsymbol{p}}_{k+1}^{+(j)}$ and its covariance matrix $P_{k+1}^{+(j)}$ have to be transmitted to FriWalk i, thus "interlacing" the two subsystems. In particular, $P_{k+1}^{+(j)}$ has to be included in the Kalman gain, as it will be shown in the following, to keep into account the fact that $\hat{\boldsymbol{p}}_{k+1}^{+(j)}$ is affected by some uncertainty.

The problem of intermittent observations in the case of a simple KF and the related stability issues are deeply analyzed in [60]. By extending a similar approach to the system at hand for each pair of agents i and j, we can define a binary random variable $\gamma_k^{(i,j)}$, which must be set equal to 1 if FriWalk i is able to observe FriWalk j at time kT_s, or 0 otherwise. Similarly, $\gamma_k^{(i,i)}$ is equal to 1 if FriWalk i is able to detect a landmark at time kT_s, or 0 otherwise. Starting from the basic update step equations of an EKF and assuming to replace the variance of real measurements with a large dummy value anytime $\gamma_k^{(i,i)}$=0, after some algebraic steps it can be shown that if the dummy variance tends to infinity [60], then the update equations of the IEKF running on agent i become

$$\begin{aligned} \hat{\boldsymbol{p}}_{k+1}^{(i)} &= \hat{\boldsymbol{p}}_{k+1}^{+(i)} + K_{k+1}^{(i)} \Gamma_{k+1}^{(i)} \left[\boldsymbol{z}_k^{(i)} - \tilde{\boldsymbol{h}}^{(i)}\left(\hat{\boldsymbol{s}}_{k+1}^{+}\right)\right] \\ P_{k+1}^{(i)} &= P_{k+1}^{+(i)} - K_{k+1}^{(i)} \Gamma_{k+1}^{(i)} \tilde{H}_{k+1}^{(i,i)} P_{k+1}^{+(i)} \end{aligned} \qquad i=1,\dots,N, \qquad (13)$$

where $\Gamma_{k+1}^{(i)}$ is a $M(N$-1)+3x$M(N$-1)+3 diagonal matrix with binary random variables only on the main diagonal, since all observations can be reasonably assumed to be independent, $\tilde{H}_{k+1}^{(i,i)}$ is the Jacobian of $\tilde{\boldsymbol{h}}^{(i)}(\cdot)$ with respect to $\boldsymbol{p}^{(i)}$ computed at $\hat{\boldsymbol{s}}_{k+1}^{+}$ and

$$K_{k+1}^{(i)} = P_{k+1}^{+(i)} + \Gamma_{k+1}^{(i)} \tilde{H}_{k+1}^{(i,i)^T} \left[\tilde{H}_{k+1}^{(i,i)} P_{k+1}^{+(i)} \tilde{H}_{k+1}^{(i,i)^T} + \tilde{S}_{k+1}^{(i)} + \tilde{R}_{k+1}^{(i)} \right]^{-1} \quad (14)$$

is the Kalman gain of the IEKF running on the i-th agent. Observe that (14) comprises two measurement covariance matrices instead of just one, i.e. the block diagonal covariance matrix $\tilde{R}_{k+1}^{(i)}$ and $\tilde{S}_{k+1}^{(i)} = \sum_{j=1 \wedge j \neq i}^{N} \tilde{H}_{k+1}^{(i,j)} P_{k+1}^{+(j)} \tilde{H}_{k+1}^{(i,j)^T}$. Matrix $\tilde{R}_{k+1}^{(i)}$ includes both the covariance matrices associated with the relative pose measurements between each pair of agents (i,j) and the covariance matrix (4) in position (i,i) due to landmark detection. Matrix $\tilde{S}_{k+1}^{(i)}$ instead takes into account the covariance matrices $P_{k+1}^{+(j)}$ of the states predicted by the agents different from i, with $\tilde{H}_{k+1}^{(i,j)}$ being the Jacobians of $\tilde{\boldsymbol{h}}^{(i)}(\cdot)$ with respect to $\boldsymbol{p}^{(j)}$ computed at $\hat{\boldsymbol{s}}_{k+1}^{+}$.
It is worth emphasizing that expressions (12) and (13) are absolutely general, but their implementation depends on the actual observation model used. With reference to the two models introduced in Section 4.3.1, in the following we will denote with subscripts A and B all the quantities which refer to *case A* and *case B*, respectively. In particular, in *case A*, $\tilde{\boldsymbol{h}}_A^{(i)}(\boldsymbol{s}_k) = \left[\boldsymbol{h}_A\left(\boldsymbol{p}_k^{(i)}, \boldsymbol{p}_k^{(1)}\right), \ldots, \boldsymbol{o}\left(\boldsymbol{p}_k^{(i)}\right), \ldots, \boldsymbol{h}_A\left(\boldsymbol{p}_k^{(i)}, \boldsymbol{p}_k^{(N)}\right)\right]^T$ where

$$\boldsymbol{h}_A\left(\boldsymbol{p}_k^{(i)}, \boldsymbol{p}_k^{(j)}\right) = \sqrt{\left(x_k^{(j)} - x_k^{(i)}\right)^2 + \left(y_k^{(j)} - y_k^{(i)}\right)^2} \quad (15)$$

And

$$\boldsymbol{o}\left(\boldsymbol{p}_k^{(i)}\right) = \begin{bmatrix} \left(x_q - x_k^{(i)}\right)\boldsymbol{\cos}\,\theta_k^{(i)} + \left(y_q - y_k^{(i)}\right)\boldsymbol{\sin}\,\theta_k^{(i)} \\ -\left(x_q - x_k^{(i)}\right)\boldsymbol{\sin}\,\theta_k^{(i)} + \left(y_q - y_k^{(i)}\right)\boldsymbol{\cos}\,\theta_k^{(i)} \\ \theta_q - \theta_k^{(i)} \end{bmatrix} \quad (16)$$

with (x_q, y_q, θ_q) being the absolute coordinates of the q-th visual landmark. Dually, *in case B* $\tilde{\boldsymbol{h}}_B^{(i)}(\boldsymbol{s}_k) = \left[\boldsymbol{h}_B\left(\boldsymbol{p}_k^{(i)}, \boldsymbol{p}_k^{(1)}\right), \ldots, \boldsymbol{o}\left(\boldsymbol{p}_k^{(i)}\right), \ldots, \boldsymbol{h}_B\left(\boldsymbol{p}_k^{(i)}, \boldsymbol{p}_k^{(N)}\right)\right]^T$ with

$$\boldsymbol{h}_B\left(\boldsymbol{p}_k^{(i)}, \boldsymbol{p}_k^{(j)}\right) = \begin{bmatrix} \left(x_k^{(j)} - x_k^{(i)}\right)\boldsymbol{\cos}\,\theta_k^{(i)} + \left(y_k^{(j)} - y_k^{(i)}\right)\boldsymbol{\sin}\,\theta_k^{(i)} \\ -\left(x_k^{(j)} - x_k^{(i)}\right)\boldsymbol{\sin}\,\theta_k^{(i)} + \left(y_k^{(j)} - y_k^{(i)}\right)\boldsymbol{\cos}\,\theta_k^{(i)} \\ \theta_k^{(j)} - \theta_k^{(i)} \end{bmatrix} \quad (17)$$

and $\boldsymbol{o}\left(\boldsymbol{p}_k^{(i)}\right)$ is the same as (16). Notice that both (15) and (16) are nonlinear, but the number of observations M is different in either case. Indeed, M = 1 in *case A* (since low-cost wireless ranging systems can hardly measure the relative orientation between transmitter and receiver with adequate accuracy), while M = 3 in *case B*.
A final remark is on communication latency, which may influence measurement uncertainty significantly, due to the difference between the time when the predicted state of agent j is sent to agent i and the moment when the relative pose of j is actually measured by i. However, if assumption 5 in Section 4.3.1 holds, then the impact of possible communication latencies can be assumed to be negligible, since the FriWalks move quite slowly (i.e. at about 1-2 m/s).

4.3.3 Simulation results

In order to evaluate the positioning accuracy with and without collaborative localization and to compare the performances of the two mutual measurement strategies described in the previous Section, the results of some meaningful Monte Carlo simulations are reported in the following. The main simulation parameters are listed below and are based on the specifications of the ACANTO project as well as on the results reported in Section 4.2.3.

- Duration of each simulated path: about 120 s;
- Number N of agents in the room: between 2 and 10;
- Room size: 100 m^2;
- Number of random paths for each FriWalk: 20;
- Robots linear velocity range: [0,2] m/s;
- Robots angular velocity range: [-π/2, π/2] rad/s;
- Sampling period: T_s = 4 ms;
- Covariance matrix of the encoders noise (based on experiments on the field): $Q_k = \mathrm{diag}(2\cdot10^{-8}\ \mathrm{m}^2;\ 2.2\cdot10^{-7}\ \mathrm{rad}^2)$;
- Reading range of the camera for landmark detection: about 1 m with an aperture angle of about 40°;
- Distance between landmarks: D=2 m;

  Covariance matrix of the measurement uncertainty associated with landmark detection (based on experiments on the field): $R_A^{(i,i)} = R_B^{(i,i)} = \mathrm{diag}(1.6\cdot10^{-3}\ \mathrm{m}^2;\ 5.0\cdot10^{-5}\ \mathrm{m}^2;\ 1.0\cdot10^{-3}\ \mathrm{rad}^2)$
- Detection range of the wireless system in *case A*: about 15 m;
- Rate of wireless distance measurements: about 25 Hz;
- Variance of the wireless distance measurement data anytime agent j is detected: $R_A^{(i,j)} = 0.45\ \mathrm{m}^2$, in accordance with [61];
- Kinect reading range (based on specifications): from 0.8 m to 3.5 m with a horizontal aperture angle of about 62°;
- Covariance matrix of the uncertainty of Kinect-based measurements anytime agent j is detected (based on experiments on the field): $R_B^{(i,j)} = \mathrm{diag}(4\cdot10^{-6}d^{(i,j)4}\ \mathrm{m}^2;\ 4\cdot10^{-6}d^{(i,j)4}\ \mathrm{m}^2;\ 3\cdot10^{-4}\ \mathrm{rad}^2)$, with $d^{(i,j)}$ being the actual Euclidean distance between agents i and j.
- Camera and Kinect (*case B*) image acquisition rate: about 10 Hz.

Figure 33 shows the average root mean square (RMS) estimation errors associated with state variables x (a), y (b) and θ (c), as a function of the number of agents moving in the room. Different bars refer to *cases A*, *B* and when no collaborative localization is used. The RMSE values of different paths for different amounts of agents are averaged over their number N. The reported results show clearly that both collaborative localization strategies enhance the accuracy in estimating (x, y). Of course, the improvement is more evident when the number of agents grows, due to both the increment of detection probability and the availability of a larger amount of measurement data from neighbors. In spite of some fluctuations due to the limited number of simulated paths, the accuracy in estimating x and y improves on average by about 0.8% in *case A* and by 3% in *case B* every time a new agent is added into the room. As far as the estimation accuracy of θ is concerned instead, results are less remarkable. First of all, the use of wireless ranging does not affect heading estimation accuracy at all, as expected, since no additional information on orientation is injected into the IEKF in *case A*. On the contrary, the Kinect-based angle measurements in *case B* improve accuracy by about 2% per agent.

In conclusion, the use of Kinects seems to be preferable for collaborative localization in the considered scenario, despite a very low duty cycle.

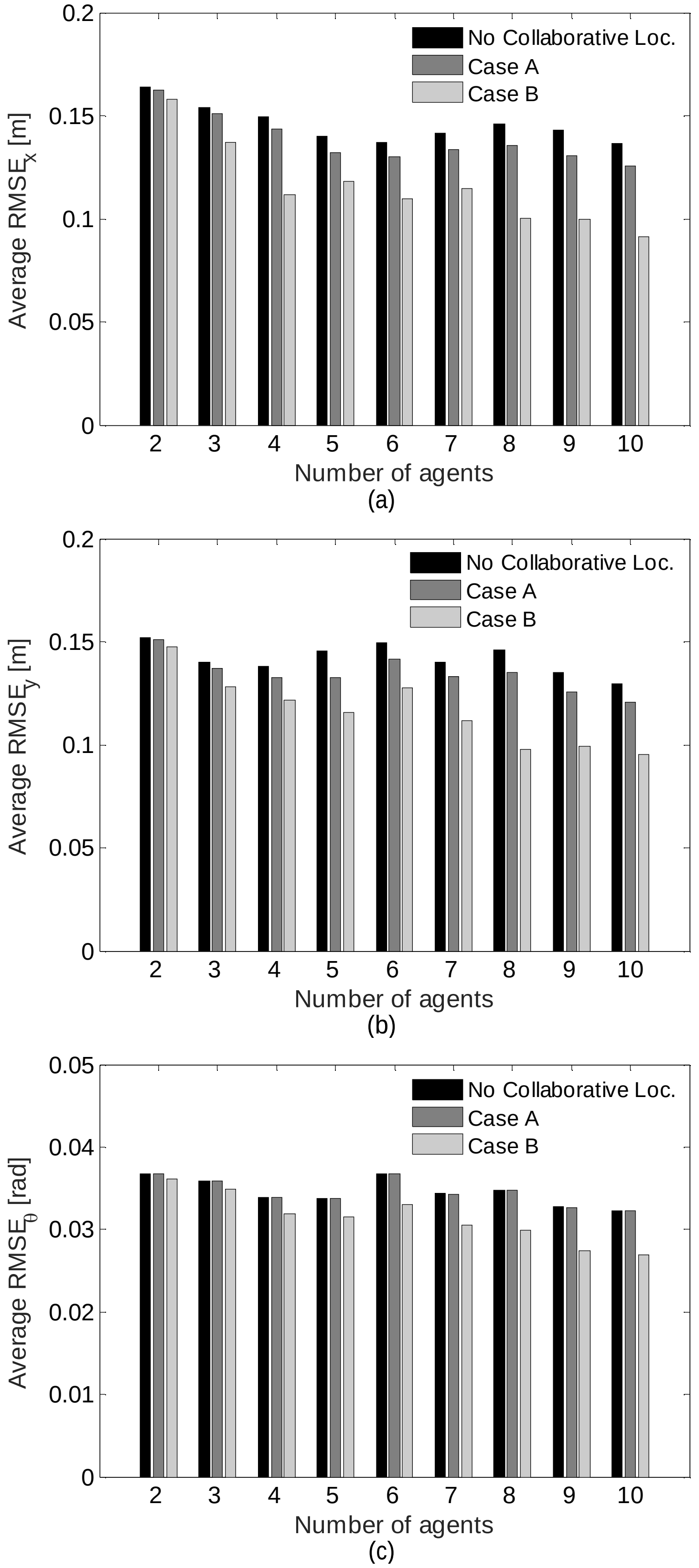


Figure 35: Average RMS estimation errors of state variables x (a), y (b) and θ (c) as a function of the number of agents present in the room, in *case A*, *case B* and when no collaborative localization is used.

Bibliography

[1] XeThru: Single-chip radar sensors with sub-mm resolution (accessed Oct 2015) https://xethru.com/
https://xethru.com/inspiration-kit.html (links to White Paper)
[2] Balakrishnan, G., Durand, F., & Guttag, J. (2013). Detecting Pulse from Head Motions in Video, 3430–3437. doi:10.1109/CVPR.2013.440
[3] Li, X., Chen, J., Zhao, G., & Pietik, M. (2014). Remote Heart Rate Measurement from Face Videos under Realistic Situations. doi:10.1109/CVPR.2014.543
[4] Mcduff, D., Member, S., Gontarek, S., & Picard, R. W. (2014). Improvements in Remote Cardiopulmonary Measurement Using a Five Band Digital Camera, 61(10), 2593–2601.
[5] Poh, M.-Z., McDuff, D. J., & Picard, R. W. (2011). Advancements in noncontact, multiparameter physiological measurements using a webcam. IEEE Transactions on Bio-Medical Engineering, 58(1), 7–11. doi:10.1109/TBME.2010.2086456
[6] Wu, H.-Y., Rubinstein, M., Shih, E., Guttag, J., Durand, F., & Freeman, W. (2012). Eulerian video magnification for revealing subtle changes in the world. ACM Transactions on Graphics, 31(4), 1–8. doi:10.1145/2185520.2185561
[7] Wang, W., Stuijk, S., & Haan, G. De. (2015). Exploiting Spatial Redundancy of Image Sensor for Motion Robust rPPG, 62(2), 415–425.
[8] De Haan, G., & Jeanne, V. (2013). Robust pulse rate from chrominance-based rPPG. IEEE Transactions on Biomedical Engineering, 60(10), 2878–2886. doi:10.1109/TBME.2013.2266196
[9] Xu, S., Sun, L., & Rohde, G. K. (2014). Robust efficient estimation of heart rate pulse from video. Biomedical Optics Express, 5(4), 1124–35. doi:10.1364/BOE.5.001124
[10] B. Bilney, M. Morris, K. Webster. “Concurrent related validity of the GAITRite® walkway system for quantification of the spatial and temporal parameters of gait”, Gait Posture 17, pp. 68-74 (2003)
[11] H.B. Menz, M.D. Latt, A. Tiedemann. “Reliability of the GAITRite® walkway system for the quantification of temporo-spatial parameters of gait in young and older people”, Gait Posture 20, pp. 20-25 (2004)
[12] Intertial Elements “Osmium MIMU22BT”, http://www.inertialelements.com/, accessed August 2015
[13] OpenShoe, http://www.openshoe.org/, accessed August 2015
[14] Vicon, 2013. Vicon: Motion capture systems. URL http://www.vicon.com
[15] Moeslund, T. B., Hilton, A., Kru, V., 2006. A Survey of Advances in Vision-based Human Motion Capture and Analysis. CVIU 104, 90-126.
[16] Poppe, R., 2007. Vision-based human motion analysis: An overview. CVIU 108 (1-2), special Issue on Vision for Human-Computer Interaction.
[17] Chen, L., Wei, H., Ferryman, J., 2013. A survey of human motion analysis using depth imagery. Pattern Recognition Letters 34 (15), 1995.
[18] Bisacco, A., Ming-Hsuan, Y., Soatto, S., 2007. Fast human pose estimation using appearance and motion via multi-dimensional boosting regression. In: IEEE CVPR.
[19] Pons-Moll, G., Leal-Taixe, L., Truong, T., Rosenhahn, B., 2011. Efficient and robust shape matching for model based human motion capture. In: Mester, R., Felsberg, M. (Eds.), Pattern Recognition. Vol. 6835 of Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 416-425.
[20] Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., Blake, A., 2011. Real-Time Human Pose Recognition in Parts from Single Depth Images.
[21] [8] Sigal, L., Isard, M., Haussecker, H., Black, M., 2012. Loose-limbed people: Estimating 3d human pose and motion using non-parametric belief propagation. IJCV 98 (1), 15-48.

[22] Sminchisescu, C., Kanaujia, A., Li, Z., Metaxas, D., 2005. Discriminative density propagation for 3d human motion estimation. In: IEEE CVPR. Vol. 1. pp. 390-397 vol. 1.
[23] Corazza, S., Mundermann, L., Gambaretto, E., Ferrigno, G., Andriac chi, T., 2010. Markerless motion capture through visual hull, articulated icp and subject specific model generation. IJCV 87 (1-2), 156-169.
[24] Deutscher, J., Reid, I., 2005. Articulated body motion capture by stochastic search. IJCV 61 (2), 185-205.
[25] Gall, J., Rosenhahn, B., Brox, T., Seidel, H.-P., 2010. Optimization and filtering for human motion capture. IJCV 87 (1-2), 75-92.
[26] Gall, J., Stoll, C., de Aguiar, E., Theobalt, C., Rosenhahn, B., Seidel, H. P., 2009. Motion capture using joint skeleton tracking and surface estimation. In: IEEE CVPR. pp. 1746-1753.
[27] Vijay, J., Trucco, E., Ivekovic, S., 2010. Markerless human articulated tracking using hierarchical particle swarm optimisation. Image and Vision Computing 28 (11), 1530-1547.
[28] Zhang, L., Sturm, J., Cremers, D., Lee, D., Oct. 2012. Real-time human motion tracking using multiple depth cameras. In: Proc. of the International Conference on Intelligent Robot Systems (IROS).
[29] J. Myers, 2003. Exercise and cardiovascular health. Circulation 107.1, e2-e5.
[30] D.R. Bassett et al., 2008. Walking, Cycling, and Obesity Rates in Europe, North America, and Australia. Journal of Physical Activity and Health, 5, 795-814.
[31] A.A. Hakim et al., 1998. Effects of Walking on Mortality among Nonsmoking retired men. The New England Journal of Medicine.
[32] J.A. Stern et al., 1994. Blink rate: a possible measure of fatigue. Human Factors: The Journal of the Human Factors and Ergonomics Society. 36.2, 285-297.
[33] J.J. Pilcher et al., 1996. Sleep quality versus sleep quantity: Relationships between sleep and measures of health, well-being and sleepiness in college students. Journal of Psychosomatic Research, 42(6), 583-596.
[34] R.S. Lazarus et al., 1962. A laboratory study of psychological stress produced by a motion picture film. Psychological Monographs: General and Applied, 76(34), 1-35.
[35] K.A. Muscatell and N.I. Eisenberger, 2012. A Social Neuroscience Perspective on Stress and Health. Social and Personality Psychology Compass, 6/12, 890-904.
[36] E. Kristal-Boneh et al., 2000. The association of resting heart rate with cardiovascular, cancer and all-cause mortality. European Heart Journal, 21, 116-124.
[37] P. Greenland et al., 1999. Resting Heart Rate is a Risk Factor for Cardiovascular and Noncardiovascular Mortality. American Journal of Epidemiology, 149(9), 853-862.
[38] R.C. Friedman et al., 1971. The intern and sleep loss. The New England Journal of medicine.
[39] P. Palatini and S. Julius, 2004. Elevated heart rate: a major risk factor for cardiovascular disease. Clinical and experimental hypertension, 26(7-8), 637-644.
[40] M. Partinen et al., 1982. Sleep disorders in relation to coronary heart disease. Acta Medica Scandinavica, 211.S660, 69-83.
[41] D. Scully et al., 1998. Physical exercise and psychological well being: a critical review. British journal of sports medicine. 32.2, 111-120.
[42] K.R. Fox, 1999. The influence of physical activity on mental well-being. Public health nutrition. 2.3a, 411-418.
[43] P. Hassmen et al., 2000. Physical exercise and psychological well-being: a population study in Finland. Preventive medicine. 30.1, 17-25.
[44] N. Sjösten and S.-L. Kivelä, 2006. The effects of physical exercise on depressive symptoms among the aged: a systematic review. International journal of geriatric psychiatry. 21.5, 410-418.

[45] D.J. Terry et al., 1993. Effects of work stress on psychological well-being and job satisfaction: The stress-buffering role of social support. Australian Journal of Psychology. 45.3, 168-175.
[46] R. Siegwart and I. R. Nourbakhsh, Introduction to Autonomous Mobile Robots. Scituate, MA, USA: Bradford Company, 2004.
[47] D. Macii, A. Boni, M. De Cecco, and D. Petri, “Tutorial 14: multisensor data fusion,” IEEE Instrumentation Measurement Magazine, vol. 11, no. 3, pp. 24–33, June 2008.
[48] H. Chung, L. Ojeda, and J. Borenstein, “Accurate mobile robot dead reckoning with a precision-calibrated fiber-optic gyroscope,” IEEE Transactions on Robotics and Automation, vol. 17, no. 1, pp. 80–84, Feb. 2001.
[49] B. Dzodzo, L. Han, X. Chen, H. Qian, and Y. Xu, “Realtime 2D code based localization for indoor robot navigation,” in Proc. IEEE International Conference on Robotics and Biomimetics (ROBIO), Shenzhen, China, Dec. 2013, pp. 486–492.
[50] G. Panahandeh and M. Jansson, “Vision-Aided Inertial Navigation Based on Ground Plane Feature Detection,” IEEE/ASME Transactions on Mechatronics, vol. 19, no. 4, pp. 1206–1215, Aug. 2014.
[51] D. Simon, Optimal state estimation: Kalman, H_∞, and nonlinear approaches. John Wiley & Sons, 2006.
[52] J. Figat and W. Kasprzak, “NAO-mark vs QR-code Recognition by NAO Robot Vision,” in Progress in Automation, Robotics and Measuring Techniques, ser. Advances in Intelligent Systems and Computing, R. Szewczyk, C. Zieliski, and M. Kaliczyska, Eds. Springer International Publishing, 2015, vol. 351, pp. 55–64.
[53] Y. Ma, S. Soatto, J. Kosecka, and S. S. Sastry, An invitation to 3-D vision: from images to geometric models. Springer Science & Business Media, 2012, vol. 26.
[54] P. Nazemzadeh, F. Moro, D. Fontanelli, D. Macii, and L. Palopoli, “Indoor Positioning of a Robotic Walking Assistant for Large Public Environments,” IEEE Trans. on Instrumentation and Measurement, pp. 1–12, 2015, available on-line.
[55] A. Colombo, D. Fontanelli, D. Macii, and L. Palopoli, “Flexible Indoor Localization and Tracking Based on a Wearable Platform and Sensor Data Fusion,” IEEE Trans. on Instrumentation and Measurement, vol. 63, no. 4, pp. 864–876, Apr. 2014.
[56] W. Li and Y. Jia, “H-infinity filtering for a class of nonlinear discrete-time systems based on unscented transform,” Signal Processing, vol. 90, no. 12, pp. 3301 – 3307, 2010.
[57] D. Fontanelli, D. Macii, P. Nazemzadeh, and L. Palopoli, “Collaborative Localization of Robotic Wheeled Walkers using Interlaced Extended Kalman Filters,” submitted for presentation to the International Instrumentation and Measurement Technology Conference (I2MTC), Taipei, Taiwan, May 2016.
[58] A. Elnashar and M. El-Saidny, “Looking at LTE in practice: A performance analysis of the LTE system based on field test results,” IEEE Vehicular Technology Magazine, vol. 8, no. 3, pp. 81–92, Sep. 2013.
[59] S. Panzieri, F. Pascucci, and R. Setola, “Multirobot localisation using interlaced extended Kalman filter,” in Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems, Beijing, China, Oct. 2006, pp. 2816–2821.
[60] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. Jordan, and S. Sastry, “Kalman filtering with intermittent observations,” IEEE Transactions on Automatic Control, vol. 49, no. 9, pp. 1453–1464, Sep. 2004.
[61] D. Macii, A. Colombo, P. Pivato, and D. Fontanelli, “A data fusion technique for wireless ranging performance improvement,” IEEE Transactions on Instrumentation and Measurement, vol. 62, no. 1, pp. 27–37, Jan. 2013.

Gait Analysis on the move: The Infinite Gait Walkway

Stefan Wakolbinger, Josef Alois Birchbauer, Severin Küberl

Siemens AG Österreich, Corporate Technology, Video Analytics Austria
{stefan.wakolbinger}{josef-alois.birchbauer}@siemens.com, severin.kueberl@student.TUGraz.at

Keywords: Gait analysis, smart walker, assistive robotics.

Abstract

In order to analyze human gait patterns, highly accurate data must be collected at high frame rates. The state of the art is to deploy a carpet-like structure instrumented with pressure sensors, which allows for measuring position, orientation and pressure of each foot at each step.
Since such gait "walkway carpets" are highly expensive[1] and also limited in length, we propose an alternative in the form of a wheeled walker equipped with a consumer depth camera. We have designed and implemented algorithms that derive the same set of parameters from the depth data as in a gait walkway system, however without the need for the physical presence of a walkway carpet. Moreover, we are able to provide additional information, due to continuous observation of the gait cycle, i.e. not only when the user steps on the ground. In order to retrieve actual foot pressure information, we use a shoe insole sensor.
Our experiments show that the system is able to collect gait relevant data with sufficient accuracy and frame rates. While the feet's position accuracy depends primarily on the noise of the depth sensor and is typically at a precision of less than 3 mm, the orientation accuracy is around 1-2 degrees for typical foot orientations.

1 Introduction

Gait analysis is the systematic study of human walking using the eye and brain of experienced observers, augmented by instrumentation for measuring body movements, body mechanics and the activity of the muscles [1]. Changes in gait reveal key information of special interest to tracking the evolution of different diseases: (a) neurological diseases such as multiple sclerosis or Parkinson's; (b) systemic diseases such as cardiopathies (in which gait is clearly affected); (c) alterations in deambulation dynamics due to sequelae from stroke and (d) diseases caused by ageing, which affect a large percentage of the population [3]. Accurate and reliable knowledge of gait characteristics at a given time, and even more importantly, monitoring and evaluating them over time, enable early diagnosis of diseases and their complications and help to find the best treatment. Continuous gait analysis can also assess the risk of falling, e.g. stride-to-stride variability has been shown to be an effective predictor of falls [2].
The traditional scales used to analyse gait parameters in clinical conditions are semi-subjective, carried out by specialists who observe the quality of a patient's gait while the patient walks. This is sometimes followed by a survey in which the patient is asked to give a subjective evaluation of the quality of their gait. A disadvantage of these methods is that they give subjective measurements, particularly concerning accuracy and precision, which in turn have a negative effect on diagnosis, follow-up and treatment. Progress in new technologies has given rise to devices and techniques that allow for objective evaluation of various gait parameters, resulting in more efficient measurement and providing specialists with a large amount of reliable information on patients' gaits. This reduces the margin of error caused by subjective techniques. Two such measurement tools commonly used in clinical gait evaluation are force platforms or gait walkways, the latter being a carpet like structure instrumented with pressure sensitive elements (sensels). One system that is now in common use is the 'GAITRite®' [8][9]. Recent advances in robotics make it possible to turn a standard assistance device, such as a walker, into an augmented device. Thus existing single shot tests can be enriched by a new set of continuously measured criteria derived from the daily use of standard assistance devices [2].

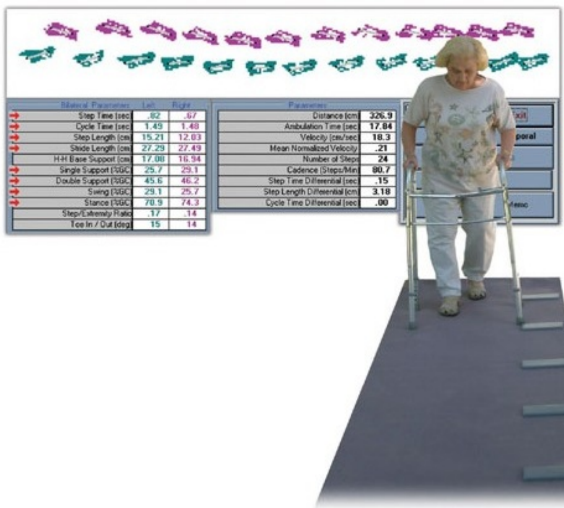


Figure 1: GAITRite® instrumented walkway system.

In this paper we propose a system that tracks specific parameters for biomechanical gait analysis. The system consists of a four-wheeled walker ("rollator") mounted with

[1] According to a desk search on various vendors between 25k€-50k€.

depth-sensors and odometers. In our work the focus is set on clinical applications and active living. Actually these two are brought together forming a continuum of data acquisition and analytics. The clinical application profits from measurements in daily life scenarios, which is likely to reduce the bias introduced in the clinical environment and vastly increases the amount of accessible data.

The main contribution of our work is the spectrum of information we derive and the "virtual" walking carpet data representation without the need for a gait walkway to be physically present. Additionally, we provide a real-time implementation (15-30 fps) that allows us to support very time-constrained algorithms.

2 Related Work

Since its beginnings in the 19th century, research on gait analysis has centred on achieving quantitative objective measurement of the different parameters that characterize gait in order to apply them to various fields such as sports, identification of people for security purposes, and medicine [3]. In our work the focus is set on clinical applications and active living. Clinical evaluation of frailty in the elderly is the first step to assess the degree of assistance they require. A comprehensive overview of the diversity and plurality of sensing modalities in the context of clinical applications is given in [3]. Reference [2] specifically raises the question if smart walkers can be used for gait monitoring and fall prevention. It considers several available smart walker implementations and concludes that standard biomechanical features such as walking speed, cadence and step length can be estimated from observing rollator walking while "*...some other information seems hard to obtained without equipping the user (3D feet positions, force pressure distribution on the ground)*". We are particularly aiming for this type of information in our work.

More specific references on individual systems in the context of an instrumented smart walker relying on a depth sensing device (e.g. Microsoft® Kinect™ sensor) are given in [4,5,6,7]. These systems do either clearly exceed our real-time runtime constraint [5,6] or do not explicitly report on the runtime behaviour, which has been a major constraint in the development of our system.

3 Parameters we aim to measure

We aim to generate all key data generated by a physical gait walkway instrumented with pressure sensels as illustrated in Figure 1, i.e. the exact placement of the feet on the ground including their orientation and the pressure force applied to the ground on foot touch. The latter actually changes over time from an initial foot contact towards the "toe-off" phase. From a representation like this a multitude of higher-level semantic information can be derived, e.g. the step and stride length, step width and the cadence. However the focus of this paper is set on generating the basic information since the derivation of the higher-level information is in most cases straightforward assuming a sufficiently precise measurement of lower-level information.

In contrast to a physical gait walkway we can also track the feet and associated key points on the feet (like the tips of the feet) continuously during the whole gait cycle, i.e. including the swing phase. This allows for e.g. temporal representations of the foot height.

Since we are by no means restricted to a straight walkway due to our Ackermann steering geometry in a front wheel steered walker we need to derive a strategy on how to visualize arbitrarily shaped – and possibly very long – walker trajectories. We opted for the following strategy:

- The chosen visualization consists in a non-length restricted but straight gait walkway. Since we observe the motion pattern of the feet from the fixed perspective of the depth camera on the walker, the walkway gets linearized implicitly. This strategy is driven by the rationale that in gait analysis as opposed to odometry, the absolute path taken is less relevant but the focus in on the relative and hence local motion pattern.
- However, an increasingly tight curve radius will affect the pattern of the feet movements. More precisely, this effect will gain the higher the change in orientation in the walker is between two subsequent steps. Instead of aiming to compensate the different curve radii of the inner and outer foot we decided just to mark the increment in the walker orientation between subsequent foot placements, which allows for information filtering in subsequent processing.

4 Our instrumented wheeled walker

Figure 2 shows the approximate depth sensor position on the walker and indicates its field of view. While this sensor is mounted on the walker, the insole sensors are worn by the user. In this work we rely on a pair of wireless "Moticon OpenGo" [10] sensor insoles, as also shown in Figure 2.

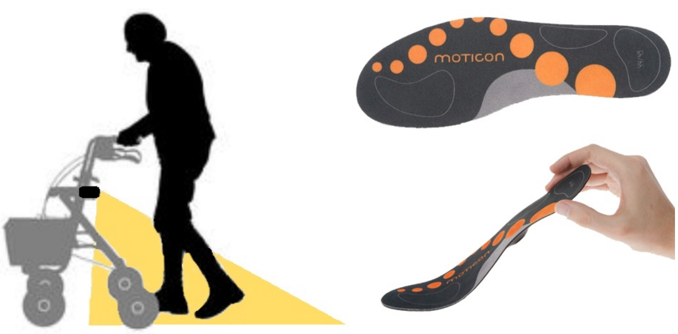


Figure 2: Instrumentation of our wheeled walker (left), utilized wireless insole "*Moticon OpenGo*" sensor (right)

Foot pressure information is mainly deemed to be relevant in clinical gait analysis while for continuous everyday inspection of gait symmetry and confidence a light-weight system without the insole sensor might also be considered a viable option, eliminating the need to instrument the user's shoes. However, the insole sensor is mandatory in order to cover the full spectrum of data of a physical gait walkway.

5 Algorithms for parameter retrieval

The algorithms designed for deriving the desired parameters solely rely on the sensor's depth data, which is represented as a depth map. An equivalent representation as a 3D point cloud can be obtained by applying

$$x = [(u - c_x)\, d]/f_x$$
$$y = [(v - c_y)\, d]/f_y$$
$$z = d,$$

where d is the depth value at pixel (u,v), f_x and f_y the sensor's focal length in pixels in x- and y direction, and (c_x, c_y) the principal point.

5.1 Ground Plane Estimation and Point Cloud Filtering

In order to allow for height estimation, the ground must be identified to provide a basis. We assume the area around the walker to be planar and apply a RANSAC-based [11] plane fitting approach to compute the plane parameters. The inlier threshold used in the RANSAC core is set to a value in the proximity of the standard deviation of the sensor noise.

Figure 3 shows the initial sensor coordinate system [x', y', z'] and the resulting ground plane coordinate system [x, y, z]. All 3D points are transformed to the plane coordinate system.

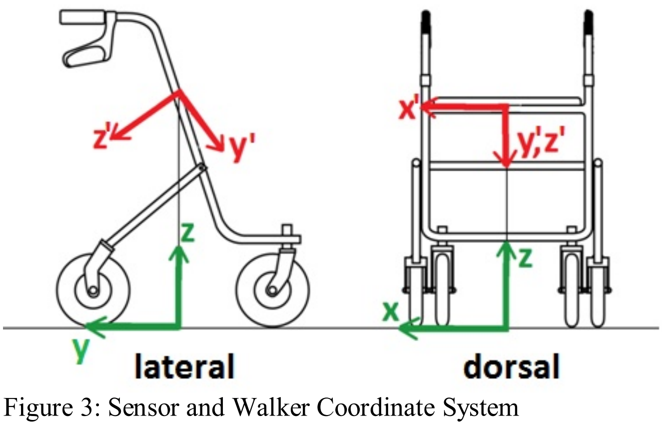


Figure 3: Sensor and Walker Coordinate System

Depending on the sensor placement, mechanical parts of the walker (e.g. the wheels) are visible in the depth map and are masked out. Since one does typically not lift a foot higher than a few centimetres during a step, 3D points located higher than 15 cm above the ground plane are also removed. Only the remaining points are used for the subsequent detection algorithms.

5.2 Foot Cluster Detection

The first step in determining the exact position and orientation of each foot is to identify the two respective clusters in the point cloud. This can be complicated by the presence of other objects on the ground that the user passes.

We use the fact that during walking, the feet are mainly oriented towards the walker and therefore visible as clusters elongated in the y-direction. All points are projected onto the ground plane, which is divided into strips in y-direction, as shown in Figure 4. In each strip, clusters in x-direction are searched (small yellow dots in Figure 4). Only points without a neighbouring, further toward the front lying point with similar x-value (~ 5 cm distance) are retained, yielding only the foremost points of each cluster (larger, green dots in Figure 4). Using this procedure, we are able to successfully identify both foot tips, even whether the feet touch (since they usually never touch at the very front).

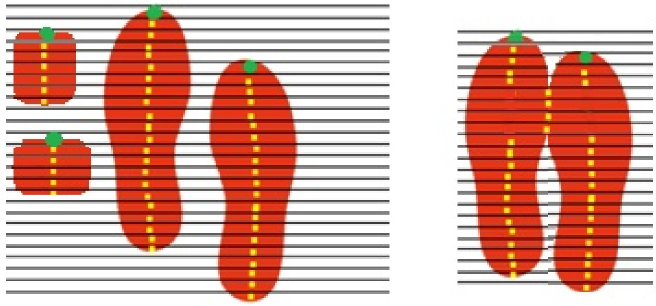


Figure 4: Foot Cluster Detection. Left: Other objects complicate the detection process. Right: Feet touch, but tips can still be identified.

In case there are more than two potential foot tips, we identify the correct ones by computing a score for each cluster and selecting the ones with the highest values.

First, all points that are within ~30 cm from the tip's y-value and differ not more than ~7 cm in x-direction are selected. The score is then computed as

$$S = N_{back} - \alpha\, N_{front},$$

where N_{back} is the number of points lying behind the tip (higher y-value), N_{front} the number of points in front and α a weight factor (set to 2 in our experiments). The score is high for foot-shaped clusters and low for small clusters, clusters that lie behind others or that are only large in x-direction.

If the scoring does not yield a clear result, we also use the position of the feet in the previous frame – if available – for correct assignment, by choosing the closest one.

5.3 Cluster Refinement

The two resulting initial foot positions indicate a rough location of the user's left and right foot tip. It is necessary to determine all corresponding 3D points for each foot, in order to enable estimation of the exact position and orientation. Especially if the feet touch (cf. Figure 4, right) this is not a trivial task. The algorithm we designed to solve this problem works as follows:

1.) Create a binary image showing all pixels in the depth map that correspond to the relevant 3D points. Find connected components and check if the initial foot positions are located in different ones. If they are, all points within the tip's component are selected for that foot.

2.) If the initial positions are within the same component, check if the feet can be separated in 3D space by performing a flood fill on the relevant part of the depth map.

3.) In case the feet are not separable using the above procedures, we seek the "best" cut through the connected region. We solve this problem directly in the depth map by defining a graph, were the cost of a connection to the neighbouring pixels is the negative depth value at these pixels. Therefore, we aim for the path with the lowest accumulated height values, which is most likely the correct cut due to the shape of the feet. The starting point for the cut is the foremost position where the feet touch. In order to make the algorithm more robust, we accumulate the depth values of several pixels in each direction before deciding which direction to move to.

Figure 5 illustrates how each of these steps can assign the corresponding points for each cluster in certain scenarios. Step 1 is successful if the data can be separated in 2D, i.e. its projection onto the ground plane. Step 2 is computationally more expensive and can perform the cluster assignment if the feet are separable in 3D space. In case the feet touch, Step 3 must be applied, which computes the ideal cut through the adjacent clusters.

Figure 5: Left: Step 1, Middle: Step 2, Right: Step 3

The advantage of this 3-step procedure is that in a typical gate cycle, in the vast majority of frames the feet can easily be separated in 2D, and the computationally more expensive subsequent steps need to be performed only when necessary. This increases the average frame rate compared to using only a single, albeit more sophisticated algorithm.

5.4 Foot Position and Orientation

In order to estimate the orientation, we perform a Principal Component Analysis on the cluster points of each foot. The direction is then set to the Eigenvector corresponding to the largest Eigenvalue of the covariance matrix.

We then project each point onto the direction vector and select the foremost point in this direction. This point, together with the direction vector and the user's foot length, unambiguously defines the foot's position.

5.5 Walker Ego-Motion

Since the coordinate system moves with the walker, it is necessary to compensate for the walker motion. The origin of the coordinate system stays at the projection of the camera centre to the ground. While the walker moves, any position recorded in the past must be moved in the opposite direction by distance the walker travelled.

One option is to use wheel odometry or inertial sensors to recover the ego-motion. In order to avoid additional hardware requirements, we implemented a vision-based method.

At each step cycle, there exists a point where both feet touch the ground. At this moment, we record the feet positions. Until this point occurs again, the walker's ego-motion can be determined by computing distance in y-Direction between the foot standing still and the stored position.

6 Experimental results

Typical output produced by our system is shown in Figures 6 and 7. Figure 6 shows that the same data is generated as in the physical gait walkway in Figure 1, i.e. the feet's position, orientation and pressure distribution at each step. While the figure only shows a short sequence, every step the user takes is visualized and the data is stored to disk for further analysis.

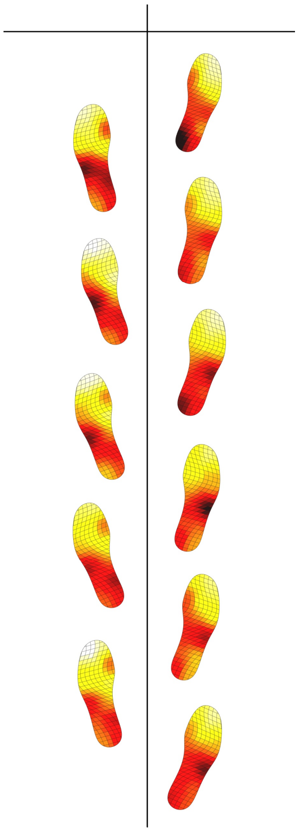


Figure 6: "Virtual Walkway" result sample.

Figure 7 shows a sample trajectory of the foot tips. It illustrates how our system is not only capable of generating data at each step on the ground, but also during the swing phase.

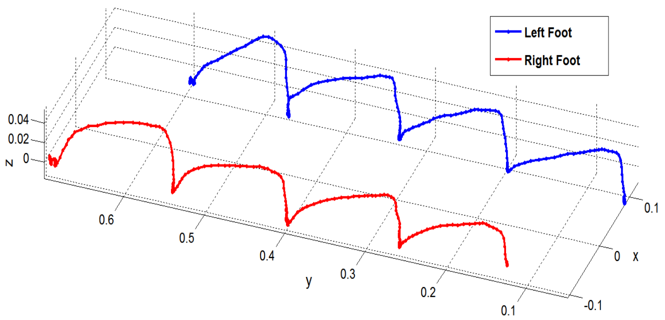


Figure 7: Sample Trajectory of the Foot Tips

The algorithms in Section 5 are designed with a strong focus on speed, which makes it possible to achieve the desired frame rate of 15-20 Hz on a single Intel®-i7 CPU core using a depth map resolution of 640x480 pixels. If higher frame rates are required, the depth map can be sub sampled to around a quarter of the resolution without influencing the results, making frame rates at around 30 Hz possible.

In order to estimate the accuracy of both position and orientation, we performed an extensive evaluation using the Microsoft® Kinect™ sensor.

For ground truth generation, we printed several identical feet patterns and placed them at different positions and angles behind the walker. Since absolute trajectories and positions are not relevant for gait analysis, but only the accuracy at each single step matters, we measure relative angles between the patterns and the distances between the foot tips.

As shown in Table 1, the average position accuracy turned out to be slightly less than 3 mm, evaluated in 20 measurements. The error is independent of the step length. Part of the deviation can be explained by the average 3D point resolution of ~1.5 mm and minor inaccuracies at ground truth capturing. Table 2 shows the results of the angle accuracy evaluation. The error increases with the angle, mainly due to occlusions. However, at typical angles when walking (0-15°) the average error of 1.6° is only slightly higher than the ground truth accuracy.

N	μ_{error}	Med_{error}	σ_{error}
20	2,96 mm	2,93 mm	1,68 mm

Table 1: Position Accuracy

Angle	N	μ_{error}	Med_{error}	σ_{error}
0°–15°	40	1,62°	1,39°	1,17°
15°-30°	40	2,26°	1,86°	2.05°
30°-45°	40	3,16°	2,86°	2,18°

Table 2: Angle Accuracy

For comparison, we have evaluated the accuracy of an inertial measurement unit (IMU), namely the *Inertial Elements Osmium MIMU22BT* [12]. Osmium produces MIMUs (*multi IMU*) that operate by fusing the measurements of several low cost sensors resulting in enhanced measurement performance. The Osmium MIMU22BT is closely related to the *OpenShoe* project, an open source foot-mounted inertial navigation system (INS) [13] initiative. We used *OpenShoe* scripts for data acquisition. While manual calibration can be performed for each individual device using a special calibration object, we used the manufacturer default calibration for practical considerations regarding a potential later deployment, i.e. for being applicable for our target group simplicity in deployment is a factor of high importance.

As shown in Figure 7, the IMU has been attached to the tip of the foot. Table 3 shows the evaluation results. Compared to our results, it turns out that the angles can be measured more accurately using the IMU, but the position error is significantly higher.

Figure 8: IMU attached to the tip of the foot.

$\mu_{Position}$	$Med_{Position}$	$\sigma_{Position}$	μ_{Angle}	Med_{Angle}	σ_{Angle}
7,3 mm	6,0 mm	6,6 mm	0,65°	0,50°	0,50°

Table 3: IMU Evaluation Results

7 Conclusion & Outlook

By upgrading a standard wheeled walker with a depth sensor (e.g. Microsoft® Kinect™), we are able to cover the same position- and orientation measurements as a deployed gait walkway instrumented with pressure sensors, which is currently the state of the art in gait analysis. In addition, we are able to produce continuous measurements during the whole gait cycle, i.e. including the swing phase. This is achieved at rates of 15-30 Hz (depending on the hardware and resolution), which allows for real-time gait pattern analysis.

While the feet's position and orientation are obtained with sufficient accuracy using the depth sensor, foot pressure measurements demand additional hardware in form of a commercialized insole sensor. Nevertheless, such a sensor is still an order of magnitude cheaper than a fully instrumented walkway system.

Depending on the user's gait pattern, occlusions can affect the system's ability to capture the feet positions. In our future work, we intend to investigate to what extent mounting a second depth sensor yielding an additional viewpoint can overcome these problems. Recent innovations in 3D depth sensing (e.g. Intel® RealSense™ R200/F200, PMD® CamBoard pico flexx) will be considered and support our aims twofold: First, the form factor/power consumption will allow for a seamless integration into the walking frame. Second, we expect that some combination of sensing devices is likely to support our aim for outdoor/sunlight compatibility as required for continuous measurements in daily activities.

Furthermore our future work will address the measurement of additional data relevant to gait analysis, e.g. position of the knees in 3D space and the angles between the lower leg and the upper leg, and the lower leg and the foot, respectively. That way we want to produce a skeletal animation of the limb movement during motion, as well as derive higher level semantic information like the joint angle and angular velocity plots discussed in [1].

Acknowledgements

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under Grant agreement n. 643644 ("ACANTO – A CyberphysicAl social NeTwOrk using robot friends"). For further information about the project see http://www.ict-acanto.eu/ .

References

[1] D. Levine, J. Richards. M.W. Whittle "Whittle's Gait Analysis", Fifth Edition (2012), p. ix

[2] C. Dune, P. Gorce, and J.-P. Merlet, "Can smart rollators be used for gait monitoring and fall prevention?", in *IEEE/RSJ Int. Conf. on Int. Rob. and Sys.*, Villa Moura, Algarve, Portugal (2012).

[3] A. Muro-de-la-Herran, B. Garcia-Zapirain, A. Mendez-Zorrilla. "Gait Analysis Methods: An Overview of Wearable and Non-Wearable Systems, Highlighting Clinical Applications", *Sensors*, Vol. 14, Pages 3362-3394, (2014)

[4] C. Joly, C. Dune, P. Gorce, P. Rives. "Feet and legs tracking using a smart rollator equipped with a Kinect". *In Workshop on "Assistance and Service Robotics in a Human Environment" in conjonction with IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS,* Tokyo, Japan, (November 2013)

[5] R. Zhi-Ling Hu, A. Hartfiel, J. Tung "3D Pose Tracking of Walker Users' Lower Limb with a Structured-LightCamera on a Moving Platform". *Computer Vision and Pattern Recognition Workshops (CVPRW)*, (2011)

[6] R. Zhi-Ling Hu, (2011). "Vision-Based Observation Models for Lower Limb 3D Tracking with a Moving Platform". *Master thesis University of Waterloo*, (2011)

[7] C. Zong, P. Bidaud, X. Clady, V. Pasqui, L. Saint-Bauzel, R.B Benosman. "A mobile 3D vision-based embedded system for robust estimation and analysis of human locomotion", *Proceedings of the Sixteenth International Conference on Climbing and Walking Robot,* vol.2008, pp.12, (2011)

[8] B. Bilney, M. Morris, K. Webster. "Concurrent related validity of the GAITRite® walkway system for quantification of the spatial and temporal parameters of gait", *Gait Posture 17,* pp. 68-74 (2003)

[9] H.B. Menz, M.D. Latt, A. Tiedemann. "Reliability of the GAITRite® walkway system for the quantification of temporo-*spatial parameters of gait in young and older* people", *Gait Posture 20,* pp. 20-25 (2004)

[10] Moticon "OpenGo" http://www.moticon.de/products/science-research, accessed August 2015

[11] M.A. Fischler, R.C. Bolles, „Random Sample Consensus: A paradigm for model fitting with applications to image analysis and automated cartography", *Communications of the ACM*, pp. 381-395 (1981)

[12] Intertial Elements "Osmium MIMU22BT" http://www.inertialelements.com/, accessed August 2015

[13] OpenShoe http://www.openshoe.org/, accessed August 2015