UNIVERSIDAD POLITÉCNICA DE
MADRID

DEVELOPMENT OF A
MULTI-MODAL SYSTEM TO
PROVIDE E-HEALTH RELATED
SERVICES BASED ON INDOOR
POSITIONING

PhD Thesis

Gustavo Hernández-Peñaloza

2019

Doctoral Thesis

Title: Development of a multi-modal system to provide e-health related services based on indoor positioning

Author: Mr. Gustavo Hernández-Peñaloza

Thesis Advisor: D. Federico Álvarez-García

Departament: Departamento de Señales, Sistemas y Radiocomunicaciones - SSR.
// Signals, Systems and Radiocommunications Department.

2019

UNIVERSIDAD POLITÉCNICA DE
MADRID

DEVELOPMENT OF A
MULTI-MODAL SYSTEM TO
PROVIDE E-HEALTH RELATED
SERVICES BASED ON INDOOR
POSITIONING

Gustavo Hernández-Peñaloza

2019

Resumen

Esta tesis doctoral está enfocada, principalmente, en la creación de modelos para la monitorización de personas en interior en sistemas de cuidado integral sostenible. Estos modelos se basan en el uso de herramientas de Tecnologías de la Información y las Comunicaciones (TIC) en combinación con técnicas avanzadas y algoritmos de procesamiento de señal. La investigación se centra en la aplicación de nuevas técnicas de clasificación y predicción sobre la información recopilada de múltiples fuentes (sistemas de sensores). Estos algoritmos se presentan en modelos de fusión de datos que se usan para crear servicios personalizados a personas mayores, y especialmente, a personas afectadas por Enfermedad de Parkinson.

Por múltiples modalidades de información se entienden las mediciones que se pueden recopilar de diversos sensores físicos desplegados en un espacio interior. Por ejemplo, los movimientos de las personas pueden ser capturados mediante sensores visuales (cámaras RGB y / o RGB-D), sin embargo, se puede fusionar con información inercial extraída de pulseras o con información recogida de dispositivos de una red de sensores inalámbricos, o incluso, de sensores binarios convenientemente desplegados (por ejemplo en las puertas para detectar salida / entrada de personas). El desarrollo de algoritmos para explotar las similitudes o patrones de comportamiento da lugar a mejoras significativas en la detección / predicción de eventos particulares a los usuarios objetivo.

Con respecto a los grupos objetivo, esta tesis se centra principalmente en pacientes con enfermedad de Parkinson y, por lo tanto, en el desarrollo de algoritmos para detectar eventos relevantes asociados con dicha enfermedad. Para lograr tal objetivo, un conjunto de características y descriptores se definen, extraen, preprocesan y modelan. Los modelos creados se basan tanto en las técnicas tradicionales de estimación estadística como en novedosas técnicas de aprendizaje profundo o Deep Learning.

Por vector de características, la investigación se refiere a la cualquier conjunto de datos que permita extraer patrones que tengan un mayor impacto en la tarea de detección, así como las estrategias de ponderación adecuadas para todas las funciones disponibles. El objetivo es mejorar el rendimiento de los algoritmos en términos de precisión.

Finalmente, la extrapolación y difusión del alcance de los algoritmos se logra mediante la creación de diversas aplicaciones que emplean los conceptos / arquitecturas basadas en los algoritmos mencionados anteriormente. En múltiples contextos, se observa cómo las técnicas de procesamiento podrían aplicarse a diversas tareas, como la clasificación por género de películas, mostrando su buen desempeño en comparación con los métodos tradicionales de descriptores visuales. Además, los métodos presentados en esta tesis también pueden aumentar la eficiencia de las redes Generativas, también conocidas como GAN de sus siglas en inglés (Generative Adversarial Networks) al acelerar el proceso de entrenamiento. Por lo tanto, el set de algoritmos y herramientas presentadas en esta tesis, métodos y modelos propuestos contribuyen al desarrollo de sistemas modulares, especialmente dedicados a la atención de la salud de las personas mayores afectadas con enfermedades del sistema nervioso, pero generalmente extensibles a otros campos.

Las hipótesis formuladas en esta tesis se validaron mediante la realización de múltiples experimentos con usuarios finales reales, creando conjuntos de datos (datasets) que pueden ser interesantes para la comunidad científica. Debido al entorno multidisciplinario en el que se desarrolló esta tesis, se han considerado aspectos metodológicos, sociales y perceptivos. En consecuencia, algunas pruebas realizadas se evaluaron mediante métodos cuantitativos (precisión, recuperación) y cualitativos (aceptación, utilidad).

Summary

This doctoral thesis is devoted to the modelling of sustainable Integrated Healthcare systems based on the use of Information and Communication Technologies (ICT) tools in combination with advanced processing algorithms. Research is focused on applying novel classification and prediction techniques to information collected from multiple sources. These algorithms jointly form fused models aimed at providing personalized care services to elderly patients.

Multiple modalities of information are understood as all the measurements that can be collected from diverse physical sensors. As an example, person movements can be captured by visual-based sensors (RGB and / or RGB-D cameras), however, it can be fused with inertial information extracted from smart-bands, or with information retrieved from Wireless Sensor Networks (WSN) Devices, or even, binary sensors conveniently allocated. The development of algorithms to exploit similarities yields to significant improvements in the detection/prediction of particular events associated to target groups.

With regard to target groups, this thesis is mainly centered on Parkinson Disease Patients, and therefore in the development of algorithms to detect relevant events associated to such disease. For this ambition, a set of features and descriptors are defined, extracted, pre-processed and modelled. The models created are built on both traditional statistical estimation techniques and Deep Learning techniques.

By significant features, research is sharpened in the extraction of data patterns that have a larger impact in the detection task, as well as the proper weighting strategies for all features available to attain a better performance in terms of accuracy, reducing the time-computing and better adapting to significant changes in the status of an environment.

Finally, extrapolation and spreading of the algorithms scope is reached by its application to diverse fields. The employment of such concepts / architectures for diverse tasks such as genre classification of movies, showing its good performance when compared to traditional visual-based descriptors methods. Moreover, methods presented in this thesis are also able to increase the efficiency of Generative Model Systems by speeding up the training process. Therefore, the toolbox of algorithms, methods and models proposed contribute to conceive modular systems, specially dedicated for elderly healthcare but generally extendable to other fields.

The hypothesis formulated in this thesis were validated by conducting multiple experiments with end users, creating datasets that can be interesting for research community. Due to the multi-disciplinary environment where this thesis took part, methodological, social and perceptive aspects have been considered. Consequently, some tests performed were evaluated via quantitative (accuracy, recall) and qualitative manners (acceptance, usefulness).

Palabras clave

Posicionamiento en interiores, Redes Inalámbricas de Sensores, Fingerprint, Filtros de Kalman, Aprendizaje profundo de máquina, Redes Generativas, Redes Convolucionales.

Keywords

Indoor Positioning, Wireless Sensor Networks, Fingerprinting, Kalman Filter, Deep Learning, Convolutional Neural Networks Convolutional Neural Networks, Generative Adversarial Networks.

Index

0.1	Integrated Healthcare Systems Context.	
0.2	Current Challenges in Integrated Health Care Systems	
0.3	Thesis Contributions	
0.4	Thesis Organization	
1	Chapter 1: Objectives and methodology	
1.1	Creation of a model based on multi-sensorial approaches to collect relevant information.	
1.1.1	Description	
1.1.2	Methodology	
1.2	Development of novel models to describe/characterize human behavior in indoor environments base on the sensorial information available	
1.2.1	Description	
1.2.2	Methodology	
1.3	Creation of a complete framework for person indoor monitoring based on multi-modal approaches, including its potential application to diverse fields of science.	
1.3.1	Description	
2	Chapter 2: Techniques and architectures of the HealthCare systems.	
2.1	Data constrains	
2.2	Ethical and Data privacy constrains	
2.3	Constrains associated to hardware /system capabilities	
3	Chapter 3: Multi-sensor Fusion Scheme to Increase Life Autonomy of Elderly People with Cognitive Problems.	
3.1	State-of-the-art	
3.2	System Architecture	
3.3	Low Level Subsystem	
3.3.1	Data Acquisition and Pre-processing (DAP)	

3.3.2	Data Synchronization and Storage	
3.3.3	Abnormal Behavior Detection (ABD)	
3.3.4	Person Identification and Multi-modal Fusion	
3.4	High Level Subsystem	
3.4.1	Electronic Health Record (EHR)	
3.4.2	Clinical Decision Support (CDS)	
3.5	Experimental results	
3.5.1	Methodology	
3.5.2	Experiments and Results	
4	Chapter 4: Indoor Positioning.	
4.1	State-of-the-Art	
4.2	Problem statement	
4.2.1	Main contributions	
4.3	Position Estimation	
4.3.1	Fingerprinting	
4.3.2	Target Position Estimation (TPE)	
4.4	Target Tracking Process (TTP)	
4.5	Multi-Sensor Processing Architectures	
4.5.1	Track-To-Track Fusion Architecture TTTF	
4.5.2	Kalman Sensor Group Fusion Architecture KSGF	
4.5.3	Alpha-Beta Filter for noise reduction	
4.6	Adaptive Fingerprinting Update	
4.7	Experimental results	
5	Chapter 5: Improving detection by using Deep Learning techniques.	
5.1	State-of-the-Art	
5.2	Problem Statement	

5.3	Fingerprint calculation	
5.3.1	RSSI Variance	
5.3.2	Wireless Technology Signal Attenuation	
5.3.3	Sampling Time per Cell and Synchronization	
5.3.4	Object or Person Tracking	
5.3.5	Orientation	
5.4	System Architecture	
5.4.1	Fingerprinting Position Estimation	
5.4.2	Gaussian Outliers Filtering and Weighted Position Estimation	
5.4.3	Step detector and velocity estimation	
5.4.4	Absolute orientation estimation	
5.4.5	Filtering Tracking with Particle Filter	
5.5	Results on Position Estimation and Tracking	
6	Chapter 6: Improving the quality of measurements estimation by using Data Augmentation.	
6.1	A Novel Approach for Indoor Tracking Using Adaptive Fingerprint Estimation Generative Models	
6.1.1	State-of-the-Art	
6.1.2	Problem Statement	
6.1.3	System Architecture	
6.1.4	Experiments	
6.2	Training speed up in Generative Adversarial Networks by Gossiping Feature Maps . . .	
6.2.1	Proposed System	
6.2.2	Experimental Results	
7	Chapter 7: Person Re-identification for Tracking Using Multi-modal Systems.	
7.1	State-of-the-Art	
7.2	Proposed method	

7.2.1	Bracelet
7.2.2	Kinect Sensor
7.2.3	Calibration
7.2.4	Feature calculation
7.3	Experimental results

8 Chapter 8: Conclusions and Future Work.

8.1	Future Lines and Dissemination results
-----	--

Annex

9 Annex2

9.1	MNIST Dataset
9.1.1	Loss Function
9.2	CIFAR-10 Dataset

Figures List

- 1 General Structure of the thesis chapters
- 2 Main Blocks of the modules developed in this thesis and its interaction.
- 3 General architecture of healthcare systems and its integration into smart cities services (High Level Subsystem).
- 4 Experiment setup for the multi-modal approach used in low level subsystem. This proposal includes Kinect, Zenith camera, WSN and eHealth bands
- 5 Description of the Zenith camera algorithm process. *from left to right in top row: (1) Original eye-fish image. (2) Camera calibration, (3) Area calibration. Bottom row, from left to right: (4) Background subtraction. (5) Blob classification. (6) Person detection. (7) Person tracking.*
- 6 Example of calibration images with chessboard pattern in different positions and orientations.
- 7 Kinect procedures description. On the left, the calibration markers are drawn. On the right side, Joint skeleton detection is shown (yellow dots) for a person walking
- 8 Example of trajectories extraction from WSN (*left*), Kinect (*center*) and Zenith camera (*right*). The red line shows the real path whereas the green line represents the estimated path. The WSN estimation accuracy is lesser than the one obtained from cameras. However, WSN coverage area is the largest of the sensors chosen. Finally, the Kinect range of detection constraints the potential applications. In the experiment, WSN is composed of 4 nodes located at the top corners of the room. Kinect sensor is located at 500,300 in horizontal and vertical axis respectively and Zenith Camera is located in the room center at ceiling.
- 9 Comparison of the estimated heading for bracelet and skeleton. The blue line is the raw data whereas the orange line represents the signal filtered. The skeleton noise is higher than the bracelet due to two main factors: (*a*) changes between 360 degrees and 0 are reflected in spite that variations are small; and (*b*) the short-term angle variation in the skeleton joints is large.
- 10 Details of high level subsystem modules and technologies integrated into a smart city infrastructure.
- 11 Example of fall detection module using fusion of sensors. In top graph, accelerometer variations for fall inference are outlined. In bottom, Kinect skeleton fall detection is drawn on the body.
- 12 Accelerometer and Gyroscope data for Freezing detection. On the (*right side*), the Freezing of Gait is appreciated whereas on the (*left side*), accelerometer data is noisy and therefore it can not be observed.

- 13 Evolution of patient location within home over a 24-hour period. The time the patient is outside is not monitored by the system.
- 14 RSSI behavior and variance along distance for different wireless sensors. WIFI variation is lesser than other technologies due to the short-term attenuation for considered distances. Conversely, XBee example is easy to fit into a valid model due to its decreasing behavior.
- 15 *(Left)* Fingerprint cells distribution used in this work *(Right)*. RSSI Histogram examples for each sensor in a particular cell. It can be observed that WiFi arises narrow range of RSSI values for the cell considered whereas bluetooth histogram probability distribution is wide spread.
- 16 Kalman Filter Algorithm steps for prediction and estimation.
- 17 Architecture patterns for sensor fusion. *(Up)*. *Track-to-track Fusion*: Separate processing stages for each sensor to achieve high-level inferences that are subsequently fused. *(Bottom) Kalman Sensor Group Fusion*: direct sensor data combination.
- 18 Main Semivariogram Patterns: *Nugget effect*: uncertainty due to lack of very close samples. *Range*: larger distances than Range are uncorrelated. *Sill*: maximum similarity value.
- 19 Top graphics represent the initial histogram representation for a cell of 50 measurements whereas bottom shows the evolution of Histograms in 300 simulations provided by AFU *(Left)* Evolution of Histogram for AFU in Bluetooth. *(Center)*. Evolution of Histogram for AFU in WiFi. *(Right)* Evolution of Histogram for AFU in 802.15.4.
- 20 *(left)* Path estimation by using Track-To-Track Fusion Architecture and *(right)* Path estimation by using Kalman Sensor Group Fusion Architecture.
- 21 Average error comparison of route estimation for every technology (no fusion applied) using Fingerprinting technique, Kernel Density Estimation and Kalman Filter.
- 22 Average error using both fusion architectures in path estimation for an experiment composed by 300 routes with 90 steps per route.
- 23 Empirical threshold for decision rule in AFU. This parameter was obtained by performing iterative simulations for 300 routes with multiple μ values in the range (0,3.5).
- 24 Error evolution for Position Estimation using AFU with a threshold $\mu = 1.5$
- 25 *Left*: This figure represents the real monitored scenario and the cells distribution for fingerprinting. Five nodes equipped with the selected wireless technologies are deployed across the area (*circles*). Doors are drawn in red and furnitures in gray respectively. *Right*: the reference coordinates system for inertial and magnetic sensors as well as the allocation of devices employed are illustrated.

- 26 Example of RSSI variance boxplots in the collected Fingerprinting database is presented for each node in different cells selected for a particular wireless technology (XBee). Variance in measurements is large, yielding to significant fluctuations in the signal distribution over a particular location (i.e. a cell in the monitored room).
- 27 An example of different orientations in the same cell is presented. The orientations have been concentrated in four angle ranges. The large variability in the measurements depends heavily on the device's orientation in each cell.
- 28 System architecture is composed by several stages: *First*, a synchronization process and a Deep Learning approach is used to perform the fingerprinting estimation (green boxes) explained in Section VA. *Second*, a Gaussian Filter is proposed to reduce outliers in the estimations (orange boxes) presented in Section VB. *Third*, the step, velocity estimation and absolute orientation (blue boxes) are introduced in Section VC and VD respectively. *Fourth*, Particle Filter tracks the person based on a realistic movement model giving the final position (yellow boxes) introduced in Section VE.
- 29 General Scheme of the Training process. *From top to bottom*: Fingerprint creation from the N Network nodes for each particular technology in addition to the orientation values (Yaw) is depicted. The length of the feature vector inserted into the Neural Network is $(I(nodes) \times \lambda(technologies)) + 1(Yaw)$
- 30 Deep Neural Network composed of: an input layer with size the selected feature vector, two hidden layers and the last layer with size equal to the number of cells with a final Softmax activation.
- 31 Example of the Gaussian Outliers Filtering. The bidimensional Gaussian is centered at the previous estimated position and the deviation is the velocity calculated at the current stage. The cell estimated by the Neural Network, in this case, is wrong so the filter applies a low probability to it. In this manner, the larger probability is assigned to the nearest cells (the ones that the filter covers).
- 32 *Left*: Three axis accelerometer measurements. Regarding the proposed referenced coordinate system it can be shown that the Y axis measurements are around $1g$ representing the gravity force ($9.8m/s^2$) in this axis. *Center*: Example of peaks/valleys detection. Red dots represent valleys and green dots represent peaks. It can be observed that the amplitude needs to be large between peaks/valleys to detect them correctly. *Right*: Estimated velocity after Monte Carlo Simulation to get the parameters for the movement model and filtering the output.
- 33 Magnetometer calibration results. *Left*: uncalibrated normalized data points, *right*: calibrated normalized data points after ellipsoid calibration.
- 34 Estimation accuracy in position estimation using several methods with different feature vectors in two main cases: The first case (*left*) corresponds to perform the normal Fingerprinting estimation based on the selected method, the second (*right*) shows the results after applying the GOF and the weighted combination proposed in this work. For horizontal axis X, B and W denote XBee, Bluetooth and WiFi technologies respectively.

35	Monte Carlo simulation over a range of possible values for adjust the the step length model to the detected peaks and valleys by the adaptive step jerk algorithm.
36	Particles (blue dots) managed by the Particle Filter at different moments during the tracking. The red dots represent the estimated position at each moment based on the particles whereas blue line depicts the final path estimated using the proposed architecture.
37	Training Loss and Accuracy for Training and Validation sets in two hidden layers (256, 64 neurons respectively) neural network configuration.
38	Images of the fingerprint measurement setup
39	General architecture of a AC GAN
40	RSSI Evolution for a static point observed from multiple Access Points for a period of 24 hours.
41	AC-GAN architecture for the Discriminator $D(\cdot)$ in the fingerprinting samples creation .
42	AC-GAN Architecture: Generative backbone $G(\cdot)$ back for fingerprinting samples creation
43	Example of the Position Estimation for a target located at cell number 10. The estimations were performed in cells 6 and 7 respectively.
44	Main evaluation metrics for the experiments proposed with the Generative Adversarial Networks: (<i>left</i>) Accuracy and (<i>right</i>) Loss
45	Main evaluation metrics for the experiments proposed with the Generative Adversarial Networks: (<i>left</i>) Accuracy and (<i>right</i>) Loss
46	Gossip Adversarial Features GAN Architecture. The dotted lines indicate that the Feature Maps from (D) in this layer are passed in a weighted manner to the Generative side (G)
47	LS-GAN results on Fashion-MNIST Dataset for epochs 5, 25 and 50 respectively
48	GAF-LSGAN results on Fashion-MNIST Dataset for epochs 5, 25 and 50 respectively .
49	LS-GAN results on Fashion-MNIST Dataset for epochs 5, 25 and 50 respectively
50	GAF-LSGAN results on Fashion-MNIST Dataset for epochs 5,25 and 50 respectively . .
51	ACGAN results on CIFAR-10 for epochs 5, 25 and 50 respectively
52	GAF-ACGAN results on CIFAR-10 Dataset for epochs 5,25 and 50 respectively
53	ACGAN results on CIFAR-10 for epochs 5, 25 and 50 respectively
54	GAF-ACGAN results on CIFAR-100 Dataset for epochs 5,25 and 50 respectively
55	GAN results on COCO dataset for epochs 10, 33 and 55 respectively

56	GAF-GAN results on COCO Dataset for epochs 10,33 and 55 respectively
57	Kinect calibration tool
58	Pitch calculation
59	Bipartite graph representing the association links of the N Bracelets with the available skeletons
60	Example of Yaw variation over time for both sensors. This case corresponds to the episode 3 described in section 7.3. On the left, the MSBand (blue) matches with the cyan skeleton. On the right hand, the other MSBand in the experiment matches with the red Skeleton. Notice that the green and cyan lines correspond to the same person whose skeleton was lost and recovered again.
61	WGAN results on Fashion-MNIST Dataset for epochs 5, 25 and 50 respectively
62	GAF-WGAN results on Fashion-MNIST Dataset for epochs 5,25 and 50 respectively
63	WGAN_GP results on Fashion-MNIST Dataset for epochs 5, 25 and 50 respectively
64	GAF-WGAN_GP results on Fashion-MNIST Dataset for epochs 5,25 and 50 respectively
65	DRAGAN results on Fashion-MNIST Dataset for epochs 5, 25 and 50 respectively
66	GAF-DRAGAN results on Fashion-MNIST Dataset for epochs 5,25 and 50 respectively
67	Loss Function for ACGAN (<i>left</i>) and GAF-ACGAN (<i>right</i>)

0.1 Integrated Healthcare Systems Context.

With an increasingly growing population in Europe, cognitive impairments as well as heart diseases are a major social and health issue. According to the WHO in a 2018 report¹ dementia, including Alzheimer's disease remains one of the biggest global public health challenges our generation is facing. Moreover, the number of people living with dementia worldwide today is estimated at 44 million, set to almost double by 2030 and is likely to rise to about 152 million by 2050. Cognitive impairment, however, is a disabling comorbidity that represents a major challenge for HealthCare Systems and it is frequent in people affected by other diseases such as Parkinson's². In the case of people who suffer Parkinson Disease, they experience issues such as: 1) loss of judgment, 2) alterations in behavior, 3) sudden mood changes and 4) difficulties in planning and organizing, which are symptoms far less known in comparison with the symptoms related to motion alteration. Alzheimer's, Parkinson and Cardiovascular diseases, mainly found in senior people as multi-morbidities have an estimated cost for the EU economy of more than EUR 196 billion a year³, and a trillion US dollars at worldwide level with forecast estimation double by 2030. Therefore, this considerable concern is moving public authorities (National Health Systems NHSs), policy makers, researchers and private businesses in joining forces to develop holistic solutions to extend autonomy of people affected by these diseases while maintaining, or even improving, their Quality of Life (QoL).

On the one side, based on recent statistics, the senior people in Europe want to live at their home. Indicatively, only a 3.3% of the population older than 65 years old live in an institutional centre and also in Europe, a 50% of people older than 80 years old, live alone, and a 35% live as a couple⁴. This clearly states that people want to live in their own homes but people with dementia, Alzheimer or Parkinson's disease have many problems to manage alone and require for care services.

Due to the complexity and severity of their diseases, these patients require substantially more resources and still have a markedly lower quality of life than most patients with just one chronic disease. Hence, there is an urge to develop new care models for management of multi-morbidity; and to make them personalized and patient-centred.

In this context, the concept of integrated healthcare is being extended as an approach enabling all parties involved in people healthcare to collaborate, communicate and exchange information to provide holistic clinical and social treatments to improve their Quality of Life (QoL)⁵. Information and Communication Technologies (ICT) can support the integrated healthcare ambition. A wide range of specific solutions has been proposed over the years to provide health services. The goal pursued is to monitor patient activities to detect anomaly events by developing tools that allow extracting patient data via physical or virtual sensors, transforming it into useful information to health professionals for treatment customization.

¹ Alzheimer's Disease International, World Alzheimer Report 2018: <https://www.alz.co.uk/research/WorldAlzheimerReport2018.pdf?2>

² Davis AA, Racette B. Parkinson disease and cognitive impairment: Five new things. *Neurol Clin Pract.* 2016;6(5):452-458. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5100708/>

³ European Heart Network, February 2017. European Cardiovascular Disease Statistics 2017 edition. Available at <http://www.ehnheart.org/images/CVD-statistics-report-August-2017.pdf>

⁴ The Health Foundation. Evidence: helping people help themselves. A review of the evidence considering whether it is worthwhile to support self-management. London: Health Foundation;2015

⁵ Mills, Susan L., et al. "Towards consensus on self-management support: the international chronic condition self-management support framework." *Health promotion international* 32.6 (2016): 942-952. <https://doi.org/10.1093/heapro/daw030>

Figure 1: General Structure of the thesis chapters

However, these solutions have multiple issues: *(a)* External factors such as financial feasibility, *(b)* regulation in terms of data protection, privacy and related concerns and *(c)* interoperability, as there exist isolated silos and therefore delay in the adoption of integrated healthcare approaches. *(d)* Internal issues in terms of hardware and processing capabilities, software efficiency and data available constrain

the performance and reliability of the information provided.

To face with the external challenges, there is an evolving change in Healthcare paradigm along with patient engagement in care decisions (also known as empowerment), in partnership with their providers and with their employer plans ⁶. The patient experiences are being extended beyond practitioner offices because of advances in integrated healthcare. Integrated healthcare has offered medical personnel with unprecedented opportunities to provide personalized care to their patients on an anywhere, anytime basis (ubiquity).

For the case of internal concerns, the development of the Internet of Things including the use of wearable devices, has also allowed for growth within the health ecosystem, enabling seamless pairing and integration with tools ubiquity of data and accessibility services that help patients receive – and be involved with – smarter and more relevant levels of care. In fact, the WHO has a “Global Strategy on People-Centred and Integrated Health Services” from 2015, which focuses on “placing people and communities at the center of health services 2016 – 2026” ⁷. Furthermore, processing techniques play a relevant role in the healthcare paradigm in the increase of results accuracy. The exponential growing of computing capabilities has permitted to apply more robust/complex techniques that take the most of information available to extract meaningful patterns of data.

The main contributions of this thesis are the knowledge generation in the field of data processing techniques for integrated healthcare systems. Especially, in the use of advanced multi-sensorial modules for behaviour analysis. The methods proposed covered systems with multiple sensors that are properly combined to detect and track events of interest. The ambition of such events is to automatically report involved members in the healthcare chain of a patient for fast attention. Also, traceability permits health professionals to support decision-making processes.

Moreover, the proposition of fusion algorithms to create models and methods able to detect abnormal behaviors provide a novel concept, using Deep Learning approaches to detect and track people in indoor environments. Additionally, the use of gossiping architectures in Generative Networks to speed up the training process presents a contribution in the literature of such networks. The partial information utilization from the Discriminative backbone to “*gossip*” the Generative side.

Additionally, the algorithms proposed for re-identification using multiple sensors combination represent a significant contribution in this extensively topic.

0.2 Current Challenges in Integrated Health Care Systems

A digitally driven market is moved forward by devices and data. Integrated healthcare providers realize this, and understood that the potential of their market relies on using the right technologies to support their customer service teams. This technology not only transforms patient experiences, but also allows organizations to operate more effectively, both in costs and efficiency. Finally, based on the extensive consortium experience, to ensure that devices are supported appropriately, their functionality and on-boarding must be addressed from the start. One of the main problems of the current Integrate Healthcare approaches is to measure the performance of integrated care.

⁶Sassi, F. and J. Hurst (2008), “The Prevention of Lifestyle-Related Chronic Diseases: an Economic Framework”, OECD Health Working Papers, No. 32, OECD Publishing, Paris

⁷<https://www.who.int/ageing/GSAP-Summary-EN.pdf>

As aforementioned, the Information and Communication Technologies (ICT)-supported Integrated Healthcare Systems field is continuously progressing and it has attracted a considerable interest of the multiple stakeholders involved (National Health Services, Administrations, Care Service Providers involving hospitals, institutions and companies providing add-value solutions, and specially, caregivers and relatives). The attempts to create a common and inclusive roadmap for the implementation of such services has derived in diverse attention focus. In this thesis, the description and solutions are focused in the aspects related to approaches for multimodal sensing and algorithms for events detection. Therefore, the successful of an integrated model for data sensing will depend on the proper data collection, pre-processing and analysis to extract the appropriate knowledge. It involves the adequate choice in modeling, algorithms and metrics to predict/detect and extrapolate results into added-value information to professionals, caregivers, and all related parties. Accordingly, from a global view, the integrated healthcare systems are composed of:

- **Low-level systems:** These systems are designed to collect information from multiple physical sensors and interfaces to provide remote-healthcare services. The goal of these systems is to collect information via sensors to detect reported to the corresponding bodies in charge of making decisions. The main ambition of Low-level systems is to extend autonomy of elderly. Further, these systems can be divided into several submodules:
 1. *Data collection and pre-processing:* this module is in charge of the sensors data gathering and storage. It is required to investigate novel ways to acquire the larger amount of meaningful data available from users subject to external constraints in terms of intrusiveness, data privacy, data integrity among others. For this reason, research lines aim at describing the available technologies and fast parallel data collection how the information is retrieved and stored. As a general concept, to define how the data gathered is going to be used to detect the events of interest.
 2. *Data processing for abnormal detection:* This module is key in the integrated models and is source of continuous innovation due to the ability to process data collected to predict events with a large accuracy.

It is important to progress in the scope of these algorithms and associated methods. There is an extensive number of research lines for (a) classical estimation (b) machine and deep learning approaches. According to the type of events, its frequency and relevance, the proposals can be classified into real-time and non-real-time (also known as offline events).

- **High-Level systems:** These systems are in charge of process low-level system inputs, Electronic Health Record (EHR), and pathway information to provide added-value services to professionals in disease monitoring and evolution and increase the knowledge in the decision-making process. This module is composed by:
 - **Multi-source Data fusion module**, which is in charge of proper combination of data to exploit similarities and extract meaningful patterns.
 - A Recommendation/prediction engine which is in charge of transform the data form medical, physical and behavioral aspects into a user-oriented (easy to understand) manner.
 - **Electronic Health Record (EHR)** (or similar) module to store the medical information in a structured. Additionally, this module comprises the mechanisms to ensure interoperability with related systems (to ensure a complete chain)

- **Communication and visualization submodule** in charge of providing secure channels to exchange information among the related stakeholders and institutions involved in the healthcare chain and the interfaces for proper visualization and interaction. The presented work is mainly focused on the low-level modules, being constantly referenced from the part of the algorithms application and assessment.

0.3 Thesis Contributions

- Improvement in the acquisition and information modeling
- Proposition of novel architecture for data augmentation
- Enhancement of the algorithms for indoor tracking and positioning in the context of Parkinson Disease.
- Development of algorithms for abnormal events detection in the context of Parkinson Disease

These contributions to the knowledge have been disseminated to scientific community through several publications in Q1 journals and conference papers as reflected across the thesis, and more explicitly, in the section *“Knowledge contribution and dissemination”* in chapter 8.

The work of this thesis is marked and associated o several National (Spanish) and European funded research projects. In concrete, the projects are:

- *ICT4LIFE*: ICT4LIFE is a three-year project co-financed under Horizon 2020, the EU Framework Programme for Research and Innovation, that kicked off earlier this year with the ambition to provide new services for integrated care employing user-friendly ICT tools, ultimately increasing patients with Parkinson’s, Alzheimer’s and other dementia and their caregivers’ quality of life and autonomy at home⁸.

0.4 Thesis Organization

This thesis is structured according to the methodology defined in the research period and it is divided into nine chapters as follows:

- **Chapter 1** exposes the general, specific and technical objectives of this thesis and describes the research methodology to reach each of these ambitions.
- **Chapter 2** provides the details on how multi-modal approaches can cover with the ambitions in the mentioned topics, being specially devoted to the technological aspects of data collection and its potential fusion for events detection.
- **Chapter 3** establishes the techniques and algorithms developed, stressing its application into real (market) systems and be adopted depending on its nature, and it provides a general architecture where these algorithms can be implemented. This chapter is devoted to describe the main features

⁸<https://ict4life.eu>

of approaches mentioned in chapter 4 and how these features can be employed to detect events and what add value can serve to high level applications.

- **Chapter 4** describes the set of algorithms proposed for event detection, starting from the information gathered by sensors (Chapter 3), and its content (Chapter 4). Chapters 3,4,5 and 6 represent the core proposition of this thesis and are aligned with the ambitions presented in the previous subsection: “Improvement in the acquisition and information modelling”, “Enhancement of the algorithms for indoor tracking and positioning” and “Development of algorithms for abnormal events detection”. These three chapters are fundamental and contain the main contributions over the State-of-the-Art and therefore have their internal division and State-of-the-Art, tests and validation respectively.
- **Chapter 5** remarks the impact of the research proposed, the roadmap to transform the applied research into realistic business opportunities and its financial viability are outlined.
- **Chapter 6** is devoted to investigate novel techniques based on Generative models to perform “*Data augmentation*” techniques over the datasets with the ambition of increasing the measurements in training and, consequently, improve the accuracy of the estimation techniques.
- **Chapter 7** provides an application of the systems presented by proposing a method for re-identification of people in an indoor environment based on the sensors information available.
- Finally, **Chapter 8** presents the conclusions and future lines of research. A summary of the progress in the knowledge that this thesis generated, as well as the potential future research topics and open branches.
- **Chapter 9** outlines the bibliography, acronyms and complementary material (annex) that were cited in the thesis but are transverse to the research performed.

1 Chapter 1: Objectives and methodology

The main goal of this thesis is **to develop a set of algorithms for indoor positioning, based on multi-modal approaches, to improve the continuous monitoring of persons in indoor environments.** The context allocates this thesis in the field of elderly care, especially people affected by Parkinson Disease PD. For this purpose, in this thesis, novel approaches for smart data acquisition from multiples sources and algorithms for prediction / classification tasks are presented. The general development of the algorithms are included into the integral healthcare paradigm and therefore, these algorithms can be integrated in such systems, according to the investigation of these systems as stated in chapter 2 of this thesis.

This general objective is concreted and distributed in three contributions that present the progress beyond the state-of-the-art of this thesis and that are outlined in sections 1.1, 1.2 and 1.3 of this chapter. In addition, each of this sections describes the methodology adopted to reach the objective, as well as specific aspects and issues that were faced in the research period for validation and demonstration of its usefulness.

These objectives were proposed as a response to the needs identified in the integrated healthcare paradigm mentioned in the previous chapter. First, it is important to design systems able to capture information of interest for the detection process. Second, to define and model the most influential parameters in the decision making process and third, to develop the algorithms that retrieve all data sources mentioned to provide the results expected. This situation is described in Figure 1, where the three main blocks are depicted and the data flow and interaction between modules is presented.

This is due to the strong relationship of the research period with the development of the H2020 project ICT4LIFE.

1.1 Creation of a model based on multi-sensorial approaches to collect relevant information.

1.1.1 Description

Integral healthcare systems, as systems in general, employ data coming from multiple sources (explicit and implicit information) that could be used to model typical behaviors. There are multiple aspects that impact the manner data is collected, including data protection, privacy, and synchronous, among others. These aspects will be further described in chapter 2, however, the objective is to capture low-level (physical) information to better support day activity of elderly in indoors.

Therefore, on the basis of multi-modality in data collection, several sensors are designed to be included in the information gathering, and it is assumed that the information collected is not sensitive and that in any case, this information will permit the “individualization” of a person. In this module, the set of technologies and its potential contribution to the detection of abnormal behaviors associated to patients is provided. The sensors involve:

- *Smartbands* that contain on-board sensors to measure biological status of a patient, and inertial sensors to model the patients movement

Figure 2: Main Blocks of the modules developed in this thesis and its interaction.

- Wireless transmitters to detect the Received Strength Signal Indicator (RSSI), which is commonly used to estimate the position of a person.

- *Binary sensors*, employed to make hard decisions (i.e. is a patient getting in /leaving the home).
- *Depth sensors*, to detect the skeleton of a person and to obtain information of their movement
- Routes extracted from visual sensors.

This information must be pre-processed to obtain comparable measurements. This is a non-trivial procedure as the sampling frequency of the mentioned devices significantly vary. Furthermore, there are specific considerations for some of the technologies mentioned, but transverse to the collection process. Additionally, there are some pre-processing tasks that will ease the features extraction procedures and that optimize the performance of the system in terms of data occupation, time consumption and processing of data.

1.1.2 Methodology

To reach a robust collection system, the following methodology was established:

1. To examine the integrity of data collected, regardless of the source (sensor).
2. To assess the synchronization strategy to safely store information that could be re-usable in a straightforward way.
3. Analyzing the State-of-the-Art related to the collection of data, the similarities and differential aspects of the proposed system.
4. To generate a complete workflow of data collection, generation and store to be used by the next modules
5. Implement the mentioned multi-sensorial system and test it directly on target users. For this purpose, it will be required:
 - Develop a multi-stage, multi-thread system able to handle thousand requests simultaneously
 - To choose what type of information will be stored for each component and what will be its future application in the detection process.
 - To validate the tool in a controlled environment
 - To validate the tool in a real environment (i.e. a target user house).
 - Once a period of data collection is finished, then remove the whole staff and perform perception tests to receive feedback
 - To compare the inputs received by the end users with the final output of the system to confirm or verify whether the modules are working properly or not.

The proper data collection is the basis of integrated healthcare systems, and therefore it is a critical aspect to face with. Associated problems such as data integrity, synchronization and security will be covered. The development of these aspects, and the validation of the technology is further described in chapter 4 of this thesis.

1.2 Development of novel models to describe/characterize human behavior in indoor environments base on the sensorial information available

1.2.1 Description

Once the sensorial data gathering process is completed, the processing modules play the relevant role in the complex task of analyzing the measurements to be transformed into meaningful high-level information. In this work, this ambition is divided into two main subtasks:

- The problem of detecting and tracking people in indoor environments and
- The detection of abnormal behaviors.

These topics are treated separately in chapters 4 and 5 respectively. The most disruptive contributions of this thesis are mainly provided in chapter 4. The indoor positioning and tracking problem is focus of extensive efforts from research community and is faced from multiple perspectives (sensorial, algorithms)... In addition, there are multiple related subtopics such as re-identification among others. As current research lines are dealing with this problem using Deep Learning techniques, research is covered in this chapter including both classical estimation techniques and Deep Learning (DL) approaches.

Therefore, once the low-level features are defined, hypothesis for proper indoor positioning and tracking are proposed. These hypothesis tackle the wide range of issues for appropriate detection, defining strategies to fuse the data available, facing with multiple sensors data and filtering outcomes.

To briefly organize the steps of these systems, we can highlight the following:

- Check data available
- Pre-process the data
- Synchronization of data intervals
- Re-identifying
- Creating reliable backgrounds (Fingerprinting)
- Estimate target position
- Filtering outliers
- Tracking

1.2.2 Methodology

The methodology adopted is stated as follows:

1. To analyze the State-of-the-Art related with the indoor positioning and tracking subject.

2. To calibrate and define dimensionality of the area under monitoring
3. To define the most relevant features from each technology for the project purpose.
4. To select the proper set of methods to be tested
5. To establish an architecture that supports the process (data forwarding, labeling, etc).
6. To implement the proposed strategies to the corresponding features and
7. To compare the results obtained.

The development of this objective and the steps followed to reach such ambition are detailed in chapter 5 of this thesis.

1.3 Creation of a complete framework for person indoor monitoring based on multi-modal approaches, including its potential application to diverse fields of science.

1.3.1 Description

Once the data is collected, parsed, processed and results are obtained, then the last stage of the set of algorithms proposed is to reach out several applications where this knowledge can be extrapolated. Taking into consideration that the algorithms developed in the previous sections are based on novel techniques (especially Deep Learning), then the scope of this section is to review diverse areas where these techniques can be applied and explore its feasibility. Of course, multiple applications require a considerable number of adaptations and, bearing the fact that we are basing the decision-making process in the information collected from diverse data sources, a complete analysis on the impact that these sources have in the final estimation is presented.

Every application has its own conditions and consequently, every algorithm applied must have its own configuration. This objective aims at covering all aspects related to configuration, setup and adaption of the algorithms with the ambition of becoming feasible solutions to be implemented in a nearby future. Therefore, the algorithms presented must allow:

- Be subject to parametrization and configuration. Although configuration is a tedious process that shall be adapted to the conditions, the most relevant parameters must be outlined.
- To be adapted to new applications with a painless procedure.

Moreover, it is important to strength the results presented in this thesis have been validated in real environments. As an example, the experiments involving human participation required special approval and the design of the disclosure agreements considering ethical, social and demographic aspects that are briefly presented in chapter 6.

2 Chapter 2: Techniques and architectures of the HealthCare systems.

There exist some strong constraints in the Clinical Decision Support System (CDS) in aspects such as security, data integrity, ubiquity among others. Moreover, there are some aspects concerning the disease dependency and specific variables and conditions of patients to be monitored. Therefore, in spite the algorithms can be general. In some cases, its application and performance can be seriously affected (or even making it not feasible) to be applied in some scenarios. In addition, elements such as data availability of computational resources/requirements can influence the performance of such systems. In this section, these constraining elements are presented, are classified depending on its nature and some intuition about where these can be critical are presented.

2.1 Data constraints

This is a fundamental constraint and limitation for most of the techniques presented in this work. Perhaps, this is one of the aspects that is gaining more relevance in nowadays research. In fact, one of the key technological cornerstones of the recent and upcoming years is data. Furthermore, the research presented in this work heavily depends on the data collected from indoor technologies in any of the scenarios considered. And going further, the availability of real data for the purpose of Parkinson Disease patients monitoring is one of the critical aspects to increase/improve the accuracy/performance of the algorithms presented in this thesis. Before going into details of the data impact, let's briefly outline the cases considered in this thesis and that will be analyzed in the next chapters:

- Create data for background (ground-truth) in indoor positioning
- Indoor positioning using deep learning techniques
- Creating data for estimation of abnormal behaviors in indoor environments

In the supposition of total lack of data for any of the algorithms proposed, the first of the aforementioned use cases is seriously affected. In fact, data, and its analytics are a trend for commercial perspectives and decision-making processes of multiple industries. Although in this scenario, it is proposed a technique to overcome the potential problems for algorithms training in indoor positioning by generating a wide number of samples, the basis of data generation (also known as data augmentation) relies on the existing data. This technique is currently being using in most of computer vision applications to increase the training datasets size. To figure out how it works, let's assume we have an image that contains an object that we want to learn to identify. We can learn some features such as color, texture and specially shape. If we want to develop a system able to detect this class of object regardless of its position, then it seems reasonable to have in the learning (training) process versions of this object allocated in diverse positions, widen, thick, rotated, even in a small reshaping, indicating the system other conditions of the same object that can also be detected. This process is called as Data augmentation and there is a vast area of research that aims at creating reliable modules that permit to generate samples for training algorithms.

The second scenario, for indoor positioning systems has special characteristics depending on the technique applied. In this thesis two types of techniques are used. The former group copes with the

classical statistics for estimation. These techniques can be sub-split into two groups: deterministic or based on ground-truth. The deterministic techniques rely on complex mathematical formulations that aim at characterizing the scenarios and propagation of signals across them. These techniques usually perform well (acceptable accuracy) in scenarios that are very controlled and are not robust to changes in the scenario conditions.

Finally, for the third scenario, the lack of data is crucial as any type of detection algorithm can be applied with no-data.

Therefore, all the mentioned cases depend on the data. The expertise gained across the development of this thesis is that data is generally hard to obtain and even more if it involves the participation of humans. Nonetheless, for the seek of practicality, this thesis was running simultaneously with a H2020 project ICT4LIFE and this coincidence allowed to extract data from real patients. In fact, the participation of the Asociación Parkinson Madrid allowed to speed up the testing process and as can be observed in the acknowledgments section, without their support, part of this work was meaningless.

2.2 Ethical and Data privacy constrains

Although the present work is mainly focused on non-intrusive data collection, there are some important limitations that must be stated. The first one, in the multimodal approach described later in this thesis, involves two types of visual sensors: the traditional RGB cameras, with the ability to record videos. These sensors are included in this thesis and were employed in the potential applications derived, however, there exists several concerns in terms of data privacy that discourage the use of this technologies. As it will be observed in subsequent chapters, this technology is partially used for multi-modal analysis, but is not the main focus of this thesis and special actions shall be taken to enable its use in healthcare paradigm. The second, is the Kinect sensor. This device has the ability of recording multiple types of data: Depth, infrared and RGB information. For the research and work related purposes, it was employed by acquiring only the skeleton (joint points) of a person.

In annex 1 of this document, the extensive research on the characteristics for each of the technologies employed and the conditions to be fully compliant with the privacy requirements are presented. However, to summarize the results:

- **Smart band:** This device is cannot be employed directly to individualize a person, however, as part of the project and research purposes, an Universal Unique Identifier UUID was associated to the hardware to be matched with the person in other high level stages.
- **Binary sensors:** No privacy issues detected
- **Wireless Sensor Network:** no privacy issues detected
- **Kinect Sensor:** It has multiple modalities. The only one that could be used to individualize persons is the RGB image. Therefore, it was not used in the thesis. Moreover, the deep image (Gray scale), was not totally clear and therefore is not employed in this project. Consequently, only the skeleton joints will be used.
- **RGB Cameras (Zenith):** This device cannot be used.

The data collection and alternatives to face with these issues are later presented in this chapter. However, bearing this type of limitations, it is important to highlight that: (a) on the one side, there is an opportunity to the integration of novel non-intrusive novel technologies. In fact, there is an extensive research effort in the fields of affective computing that employ a combination of visual sensors with other technologies and can broad the scope of these systems. (b) The data collection affects the rest of high-level systems. The nature of data gathered and its model can infer on the precision of the outputs and therefore, reliability of the system as a whole. However, regardless of the devices available, there are a set of features that are going to be presented, and that in general can be employed for person positioning, tracking and events detection in multiple scenarios.

2.3 Constrains associated to hardware /system capabilities

The third type of constrains are defined by the functional and technical requirements of the system. These requirements can condition the proper system operation and make some approaches unfeasible. Having in consideration the healthcare systems as a pyramid where the bottom (basis) is composed of a central collector (generally a PC) connected to a set of sensing devices. These devices pre-process data locally, detecting some events in real-time and being interconnected to cloud services where the entire processing of the information will be performed.

The low-level system collector needs to have computing capabilities to capture simultaneously information from the mentioned devices. Additionally, the preprocessing stages must be performed and some events shall be detected directly on this machine. The most common strategy is to balance the processing tasks to guarantee system stability. In this section, the details of the systems employed are provided.

3 Chapter 3: Multi-sensor Fusion Scheme to Increase Life Autonomy of Elderly People with Cognitive Problems.

Chapter Summary

This chapter introduces the systems proposed, its main components, the general architecture, and several low-level submodules. The main contributions for this thesis are given by the general organization of the modules, and mainly on the system functionalities implemented: The data collection modules, specially the *synchronization* for *re-identification* that will be further developed in next chapters, the *person detection and tracking* that will also be extended in the next chapters. The contents of this section are mainly associated to a Q1 Journal “*A Multi-Sensor Fusion Scheme to Increase Life Autonomy of Elderly People With Cognitive Problems,*” in IEEE Access, vol. 6, pp. 12775-12789, 2018 [37].

3.1 State-of-the-art

There is a growing interest in the employment of Information and Communication Technologies (ICT) as a response to the healthcare system requirements for aging people [3, 4, 5]. Most of the applications are mainly focused on the active homes monitoring which represents a considerable advance as the patients autonomy can be extended [3, 6]. Additionally, the multi-disciplinary health teams can obtain a continuous vital signs and behavior tracking which yields to appropriate diagnosis or treatment [8].

However, to guarantee the sustainability of these systems subject to special needs of elderly in terms of accessibility and mobility, several issues must be faced in organizational, medical, stakeholders and infrastructural aspects [8, 9].

Furthermore, elder people activities must not be only constrained to home care, but it must be considered as part of a global monitoring in their society [8, 10]. In this sense, smart cities play a key role, as some activities can be included within the range of services aiming for improving the citizens' quality of life [10]. The impact that healthcare services for smart cities can have into population has been quantified in up to 12 billion in 2020 [9]. In the context of integral healthcare applications, multiple opportunities from a research perspective arise.

In the recent years, advances in ICTs have allowed the development of low-cost devices for monitoring complex activities. As an example, camera deployments for people tracking [85], Wireless Sensor Networks (WSN) based technologies permit non-intrusive monitoring [6]. Additionally, inclusion of e-health sensors in wearable devices have increased the potential of full-time monitoring applications [16]. However, there exist several constrains associated to each technology, such as occlusions and lighting changes in camera-based systems; accuracy in wireless devices among others [17]. To tackle these factors, an interesting alternative is given by fusing the information from multiple sensors.

Therefore, a multi-modal sensor fusion scheme for health monitoring is presented. This system considers the medical, ethical and functional issues for data retrieving and synchronization in Parkinson Disease (PD) patients. Additionally, the system is intended to detect abnormal events by the proper fusion of the measurements gathered from the sensors. The system is able to detect several persons simultaneously and associate the data from the sensors considered to an individual. This is a signif-

icant advantage of the proposed system as it allows to extend the scope of detection and tracking to multiple patients.

Several works have been presented for integrated healthcare using ICT-based solutions. A complete review of Wireless Sensor Networks (WSN) systems for healthcare applications is drawn in [5]. Moreover, traditional health monitoring systems rely on the use of vision sensors [85]. Recently, implementation of depth sensors have permitted to analyze body movements in a detailed manner by the employment of interesting features such as the joint skeleton detection [7].

In [4], a complete architecture for home-care monitoring is described. In this architecture, information from infrared sensors, microphones and depth cameras is centralized in a gateway for home-care. However, the system presented in this work aims to be modular as it is able to detect events even with partial information. Furthermore, healthcare services are intended to be deployed and work in several scenarios. The system proposed is capable of monitoring several pedestrians and individualize the events detected for analysis which is called person re-identification.

The main contributions of this section are:

- A multi-sensor modular scheme for indoor people monitoring.
- A set of modules for the implementation of tracking and sensing functionalities for detection of abnormal events in PD patients.
- A framework to associate measurements gathered from diverse sensors to the corresponding person.
- A collaborative system that allows the appropriate fusion of multiple sensors.

The remainder of this section is organized as follows: in subsection 3.2, the general architecture of the proposed system is outlined. Furthermore, in subsection 3.3 and 3.4, the main modules of the proposed system are presented. In subsection 3.5, the methodological aspects of the experiments, as well as setup and results of the described modules are drawn.

3.2 System Architecture

From a global perspective, the inclusion of healthcare systems into smart cities services must address several issues, specially in terms of data privacy [5]. A robust codification process must be performed in order to guarantee that patients identity is preserved. Consequently, the overall architecture of the proposed system is illustrated in Figure 3. This architecture is mainly composed of a low level and a high level subsystem. The former is mainly devoted to the data acquisition and low level sensing whereas the latter is in charge of the data processing and inference extraction. This modular architecture involves the physical implementation of sensor systems in the places to be monitored. The aim of this system is to be integrated into a smart city in a Software as a Service (SaaS) distribution model [13]. Additionally, the high level subsystem can access information from the smart cities for decision-making procedures using a Platform as a Service (PaaS) model [15].

The low level layer is formed by (a) the Data Acquisition and Processing (DAP) module, (b) the Data Synchronization and Storage stage and (c) the Abnormal Behavior Detection (ABD) module. DAP is

Figure 3: General architecture of healthcare systems and its integration into smart cities services (High Level Subsystem).

responsible for the proper acquisition and filtering of the information gathered from diverse sensors. Further, Data Synchronization and Storage stage aims to securely save data as well as synchronization tasks for later processing. Finally, methods for the detection of events associated to people with cognitive problems are implemented in ABD. These methods rely on fused information from the data gathered in previous modules.

The high level Subsystem consists of: (a) an Electronic Health Record (EHR), (b) a Clinical Decision Support System (CDS) and (c) an authentication component. EHR is the module where the medical related information is stored. CDS contains the recommendation engine to support the decision making of health professionals. Finally, the authentication module endows the system with security layers by adding headers to protect the identity of users. In next subsection, details on the architecture subsystems are provided. This work is part of a research project aiming to develop a system to extend life autonomy of elderly people with Parkinson Disease. The experiments as methodological aspects were defined in collaboration with the Asociación Parkinson Madrid (APM) ⁹.

The novel concept is given by the low level Subsystem modules and functionalities for PDs patients care as well as its implementation. Therefore these aspects will be detailed whereas the high level modules will be briefly addressed and reviewed. The remainder module of this architecture comprises user interfaces that are out of the scope of this work.

⁹Asociación Parkinson Madrid <https://www.parkinsonmadrid.org/>

Figure 4: Experiment setup for the multi-modal approach used in low level subsystem. This proposal includes Kinect, Zenith camera, WSN and eHealth bands

Table 1: Data extracted from the selected devices. Zenith camera coverage depends on the lens employed. Kinect default resolution is 640×480 at 30 Frames Per Second (FPS).

Device	Sensor	Feature
Bracelets and Belts	Accelerometer	ax,ay,az
	Gyroscope	gx,gy,gz
	Bio-sensors	Heart Rate, Temperature
WSN	IEEE802.15.1	RSSI (dBm)
	IEEE802.15.4	RSSI (dBm)
Depth Sensor	Skeleton joints	$6 * 25 * x, y, z$
	RGB Image	up to 30 fps at 1280×1024
	Infrared	black and white object movement
Zenith Camera	RGB Image (up to 360°)	up to 1280×960 at 15 fps

3.3 Low Level Subsystem

The low level subsystem comprises the modules for data acquisition and processing algorithms to extract the information used for ABD. It is divided into three modules. The first includes data collection from diverse sensors and devices in addition to some brief pre-processing tasks (DAP). The second stage involves the data synchronization, labeling and storage. In the third stage, several processing algorithms for synchronization, association and data fusion are performed in two senses: (a) to reduce the amount of data sent to higher layers by sending only processed information and (b) to allow the system to make real-time decisions by containing the whole set of algorithms that permit the extraction of useful information from sensors. In a global overview, the multi-sensor approach of the low level subsystem for elderly people monitoring is presented in Figure 4.

3.3.1 Data Acquisition and Pre-processing (DAP)

The devices used in the system retrieve data from several sensors to extract information about elderly people activity and behavior. This data must be collected and stored before applying processing algorithms to make inferences or generate alarms yielding to useful reports for the medical teams. A high synchronization level is required to label and classify the entire dataset. Sensors deployed over diverse scenarios provide multiple data types, sampling rate, resolution and accuracy. An overview of devices and features extracted is depicted in Table 1. Specifically, the following sensors are considered:

Depth Sensor

Kinect sensor V2 is a multi-modality device developed by Microsoft [19]. It is able to provide up to 30 frames per seconds (fps) for several types of data formats as described in Table 1. Both images and skeleton joint points are labeled with the corresponding timestamps in the DAP module in the same manner as other sensors. The 640×480 resolution at 30 *frames per second* (fps) was chosen in the scope of this project. Data retrieved from this sensor will be employed for several behavioral (daily motion, etc...) and physical related problems (Festination, Freezing).

Table 2: A range of abnormal events to be detected with the corresponding functions that must be implemented to reach it. In left column the sensors involved in the multi-modal detection are described.

ABD	Functionality	Sensors
Daily Motion	Trajectories	WSN, Kinect, Zenith
Falls detection	Acceleration Bracelets and Belts	Bracelets, Belts and Kinect angular speed
Parkinson Freezing	Movement assessment	Bracelets, belts and cameras
Parkinson Festination	Movement assessment	Bracelets, belts, cameras and WSN.
Dream motion	Movement assessment Patient leaving the home	Bracelets and WSN. WSN and camera.

Zenith Camera

Traditional vision sensors have played an important role in general person detection and tracking applications. Additionally, the use of special lens can increase the range of detection which represents an important advantage for large range applications. In the scope of this work, a fish-eye camera by Vivotek [12] with a 1280×960 and a sampling of 15 *fps* will be employed for both person identification and tracking.

Wireless Sensor Networks (WSN)

A WSN consists of a group of self-organized sensing nodes deployed across the area to be monitored. There is a wide range of applications where these devices have been used [6, 5]. According to Table 2, the main goal of these devices is the daily tracking of PD patients. For this purpose, a set $\mathbf{v} = \{v_{i,\lambda} \mid i = 1, \dots, I; \lambda = 1, \dots, \Lambda\}$ of Access Points (AP) located at Cartesian coordinates $\mathbf{p}_i = (x_i, y_i) \in \mathbb{R}^2$ is defined. Furthermore, let Λ be the number of sensors that each node is equipped with. In this work, two sensors from mentioned wireless technologies are considered: IEEE802.15.1/Bluetooth v4.0 [119], IEEE802.15.4/XBee [21]. Network nodes are able to obtain the Received Signal Strength Indicator, denoted by *RSSI*, from the monitored target. Consequently, $RSSI_{i,\lambda}$ describes the RSSI measurements gathered by node i using technology λ .

Health Bracelets and Belts

Several smart devices were chosen for the testing purposes. Firstly, a commercial health band, developed by Microsoft, with a sampling rate of up to 45 samples per second has been adopted [18]. This bracelet is equipped with accelerometer, gyroscope, Galvanic Skin Response (GSR), Heart Rate (HR), skin temperature sensor and bluetooth transmitter. Furthermore, another commercial eHealth band endowed with the aforementioned sensors and magnetometer has been also adapted to the system [20]. Additionally, this band is IEEE802.15.4 XBee/Zigbee compatible [21]. The reason to evaluate both bands is to perform multiple tasks taking advantage of the wireless monitoring capabilities that WSNs are able to estimate.

In concrete, IEEE802.15.4 sampling rate is up to 30 messages per second. This information is gathered

by the WSN for position estimation. Moreover, sampling rate for Bluetooth RSSI extraction is 2 seconds. The DAP module is in charge of receiving measurements from bracelets (to obtain HR, temperature and GSR measurements) and the Wireless Sensor Networks (WSN) (RSSI from the nodes deployed). The measurements are labeled with both timestamps from bracelet devices and local synchronization of the gateway node where samples are collected.

3.3.2 Data Synchronization and Storage

Because of the nature of sensors employed, a synchronization stage is required in order to allow proper data treatment. Accuracy of timestamps labeling guarantees the similarity of measures to be analyzed. For this reason, this module will simultaneously add the common timestamps to each measurement. Additionally, the multiple sampling rates of devices considered can yield to comparison errors. Therefore, time-windows are created to process the data from multiple sensors.

The first step is to quantify the delay introduced by every sensor. Empirical experiments allowed to obtain the Zenith delay by performing an action (e.g. applause) and calculating the lag between the action realized and the processed timestamps (stored). 100 repetitions of this process yield as result that the camera delay is 2.6 seconds.

Additionally, images from cameras (RGB) are processed to extract the useful features and discarded to guarantee privacy protection. Data extracted will remain stored until the ABD access this information. Once data and events are forwarded to the high level subsystem, the local information is deleted.

Moreover, in order to ensure data privacy, a policy for storage in a non-human readable format is adopted. This policy involves a serialization process into binary files. The subsequent module (ABD), will initially perform a de-serialization task in order to extract the data.

Kinect delay depends on the move speed, therefore for slow-motion events it can be neglected whereas for fast moves, the delay is up to 200ms [44].

Finally, delays introduced in bluetooth transmission have been extensively studied [46]. In concrete, for the devices employed in this work, it has been proved to be 200ms

The second step is to define a reference for synchronization. In this work, the reference time-system is defined by using the Kinect sensor frame rate. Accordingly, the samples of all sensor devices must satisfy the following condition:

$$ts_{B_i} - \frac{1}{v_{fps}} \leq ts_{K_j} < ts_{B_i} + \frac{1}{v_{fps}} \begin{cases} i = 1, \dots, N; \\ j = 0, \dots, 5; \end{cases}$$

where ts_{B_i} and ts_{K_j} denote the band and Kinect timestamp respectively. Furthermore, the v_{fps} parameter refers to the Kinect frame rate, which is obtained by counting the number of frames per second to be processed. This condition is applied for inertial data of the N devices available and the skeletons detected (in the range 0-5) by the depth sensor.

Moreover, a smoothing filter is applied to the information gathered from all sensors to minimize the impact of outliers and some abnormal moves. In concrete, the well known Butterworth filter is applied [45]. This is a low-pass filter that aims to reduce high frequency variations.

3.3.3 Abnormal Behavior Detection (ABD)

In this subsection, the main activities to be detected as well as the multi-sensor scheme and algorithms developed are described. Firstly, it is important to highlight that this work is part of a research project which aiming is to extend the autonomy of elderly people with cognitive problems (see Acknowledgments). Therefore, general activities such as daily motion and fall detection are combined with specific targeted activities for people suffering from PD as shown in Table 2. Appropriate fusion of sensors involved will increase the accuracy of ABD.

In the next subsections, details on the functions implementation as well as the ABD decision making will be provided.

Daily Motion

Based on data from previous modules, processing algorithms are mainly focused on the fusion of sensors data to track PD patients. Non-intrusive tracking is an emerging research topic whose goal is to achieve the best accuracy in the final estimation. In this project, three sensors for tracking algorithms are considered using both wireless signals and cameras. The first system uses WSN ability to obtain the RSSI parameter from Bluetooth and XBee transmitters. By the application of advance processing techniques it is possible to estimate the position. The second system uses RGB image information to track people based on the detection of movement in the image. The third functionality is implemented by the extraction of the joint skeleton positions and estimate the orientation and position of the object sensed.

Wireless Sensor Network Tracking

WSNs represent a valuable option for multiple-person positioning and tracking due to its ability to perform non-intrusive detection and overcome the problem of visual occlusions. Furthermore, WSN does not require additional hardware to wireless transmitters for RSSI parameter detection. However, the large variance of these measurements raises an important challenge for the accurate estimation process. For the proposed system, several nodes have been deployed along scenarios considered to gather the RSSI parameter from bracelets IEEE802.15.1Bluetooth and IEEE802.15.4XBee devices.

Once the RSSI is collected by all nodes, the positioning algorithm can be applied and afterward, a filtering stage to smooth the estimation is carried out to calculate the final trajectory. In a general statement, the algorithm is composed by the following steps:

- *Fingerprinting calibration.* This technique performs an initial RSSI dataset by collecting data from bracelet wireless sensors at certain predefined locations. These locations are obtained by splitting the room into several cells [23, 24, 25].
- *Fingerprinting Kernel density Estimation.* This algorithm selects the cell that fits better between the fingerprint dataset and the current measure [26, 27, 28].
- *Orientation estimation.* In this stage, the heading angle to estimate the device orientation in the 360 degrees is obtained [61].

Figure 5: Description of the Zenith camera algorithm process. *from left to right in top row: (1) Original eye-fish image. (2) Camera calibration, (3) Area calibration. Bottom row, from left to right: (4) Background subtraction. (5) Blob classification. (6) Person detection. (7) Person tracking.*

- *Velocity estimation.* Fusing information from accelerometers and positions estimated, the person velocity can be achieved by using this algorithm [29].
- *Particle Filter Tracking.* Filtering of the estimated positions to determine the final trajectory applying a real movement model [30].

Fingerprinting is a technique commonly used to characterize the signal propagation for indoor environments [48]. Due to the hard conditions, reflection and obstacles, it is difficult to fit the propagation attenuation into a valid model. Fingerprinting tackles this problem by creating a RSSI datasets which are obtained by splitting the scenario into a set $\mathbf{c} = \{\mathbf{p}_l | l = 1, \dots, L\}$ of cells with geometrical center at $\mathbf{p}_l = (x_l, y_l) \in \mathbb{R}^2$ and collecting RSSI values for a predefined period in each cell c_l .

Moreover, taking into account the fingerprint dataset, *Kernel Density Estimation* is applied to estimate a position in the scenario. The algorithm employed in this work is called Nadaraya- Watson Kernel Regression [28]. This algorithm estimates the position by comparing the similarity degree of the fingerprinting dataset with the WSN RSSI measurements. The position $\hat{\mathbf{p}}$ is obtained by using the following equation:

$$\hat{\mathbf{p}} = \sum_{c=1}^N w_c \mathbf{p}_c \quad (1)$$

where w_c denotes the weights, with its corresponding center position $\mathbf{p}_c$. These weights are obtained by using Gaussian Kernel functions [59] as follows:

$$w_c = \frac{K\left(\frac{\mathbf{RSSI} - \mathbb{E}[\mathbf{RSSI}_c]}{h}\right)}{\sum_{j=1}^L K\left(\frac{\mathbf{RSSI} - \mathbb{E}[\mathbf{RSSI}_j]}{h}\right)} \quad (2)$$

where $K(\cdot)$ is the Kernel function employed, L is the number of total cells, $\mathbf{RSSI}$ is the vector containing current RSSI measures from WSN, $\mathbb{E}[\mathbf{RSSI}_c]$ denotes the vector with mean RSSI values from fingerprinting dataset and h is the bandwidth parameter for the Kernel function used. Further details as well as examples and public datasets of this process can be found in [38]

Particle Filter is employed for tracking purposes by applying the following movement model:

$$\begin{aligned} x_k &= x_{k-1} + v_{x_k} \cos(\theta) dt \\ y_k &= y_{k-1} + v_{y_k} \sin(\theta) dt \end{aligned} \quad (3)$$

where $\mathbf{p}_k = (x_k, y_k)$, v_{x_k} and v_{y_k} denoted the x, y axis velocity respectively, whereas θ represents the orientation. To use this model, the mentioned parameters (x, y, θ and v) must be previously estimated. The latter can be calculated from accelerometer, gyroscope and magnetometer sensors. The orientation angle for the movement model is known in the same manner as the heading angle or yaw in navigation and it can be obtained by applying the following equation:

$$\theta = \arctan\left(\frac{2q_1q_4 + q_2q_3}{2q_1^2 + 2q_2^2 - 1}\right) \quad (4)$$

where q_x are the values of the quaternion vector obtained from accelerometer, gyroscope and magnetometer sensors using Magdwick Filter [137]. This filter is able to fuse sensor measures to calculate the quaternion vector that is a orientation representation using complex numbers.

The other parameter needed is the velocity of the monitored people. This information can be estimated using accelerometer measurements. The three axis values can be fused into a single value that provides information of the acceleration behavior as a function over time:

$$a = \sqrt{a_x^2 + a_y^2 + a_z^2} \quad (5)$$

Using an adaptive threshold peak detector algorithm [31], the maximum and minimum peaks of this curve can be obtained. In addition, Weinberg formula [29] can be employed to calculate the step length of people walking. With this information, the velocity by time interval is also estimated as follows:

$$Step = v * dt = K_1 \sqrt[4]{a_{max} - a_{min}} \quad (6)$$

Figure 6: Example of calibration images with chessboard pattern in different positions and orientations.

where $0.3 < K_1 < 0.7$ is a constant value that depends on the patient height, age and a_{max}, a_{min} are the maximum and minimum values obtained from Equation (5). Once these parameters are computed, the particle filter can be utilized to iteratively estimate the position with a Kernel-based algorithm. The particle filter adopts Monte Carlo methodology to find the solution of filtering problems. The steps of the mentioned algorithm are:

- Generate P random particles with a Gaussian distribution across the considered scenario.
- Predict the new state based on the movement model selected (4).
- Calculate the weights according to the distance of the particles to the prediction.
- Re-sampling particles.
- Estimate the new state for the system.

This algorithm will return the trajectory of individuals wearing bracelets around the room. The final estimation presents high accuracy in terms of Minimum Squared Error (MSE) in comparison to the real path. Further details can be found in Section 3.5.

Figure 7: Kinect procedures description. On the left, the calibration markers are drawn. On the right side, Joint skeleton detection is shown (yellow dots) for a person walking

Zenith Camera Tracking

Zenith cameras have gained relevance as suitable option for large area tracking systems. The main advantage of these devices is given by the comprehensive detection range as they offer the widest-field-of-view possible. Computer vision algorithms are employed to detect movement on the frame and decide whether the objects detected correspond to people or not [32]. Moreover, a tracking algorithm is used over these detected persons for trajectories estimation and finally, the conversion of pixel images to real world coordinates is carried out. The entire procedure for person detection and tracking using Zenith cameras is described as follows:

- *Camera calibration.* This process is performed by using a chessboard pattern to obtain the intrinsic, distortion and extrinsic parameters of the camera chosen.
- *Area calibration.* Selection of several points on the floor to get the parameters for pixel to world coordinates conversion.
- *Histogram of Oriented Gradients training.* In this stage, the Algorithm training to detect people in the image is carried out [32].
- *Undistorted images.* Procedure executed to eliminate the omni-directional distortion effects introduced by the 360 degrees lens [33].
- *Image movement detection.* This algorithm is utilized for background subtraction of the image; giving as a result the possible motion objects in the image.
- *Blob classification.* Based on the probable motion objects detected, applying this algorithm allows to verify if this object is a person or not [65].
- *Kalman Filter Tracking.* Implement the well known Kalman Filter tracking to persons detected along the images collected [34].
- *Detection Assignment.* To allocate the most probable distribution of trackers along the images.
- *Pixels to real world coordinates conversion.* Transform the tracked pixels route to real world coordinates in the reference system.

The entire process is drawn in Figure 5, where an image of every step of the Zenith sensor is shown.

Camera calibration. This process is required in case the camera intrinsic and extrinsic parameters are unknown. It consists in obtaining several images from a chessboard pattern to detect the corners in different positions and orientations. An example of the mentioned procedure is illustrated in Figure 6. Harris corner detector is employed to extract the points from the image and therefore to estimate the intrinsic, distortion and extrinsic camera parameters using Zhengyou Zhang calibration [157].

Once the camera calibration is performed, *the area under monitoring must be calibrated.* This process allows to convert from pixels to real spatial coordinates. To carry out this task, several known points on the floor must be marked and a camera image must be taken. The calibration algorithm consists of selecting the pixel points where the real world coordinate markers are.

The next step is to train the algorithm for people detection. In this work, the Histogram of Oriented Gradients (HOG) is chosen [65]. HOG algorithm is a feature descriptor that counts occurrences of

gradient orientation in localized portions of an image. This algorithm relies on the training of a Supported Vector Machine (SVM) with images of persons across the area under monitoring. As a result, a hard decision about the objects detected in the crop image as persons is made.

In order to mitigate the effects of eye-fish lens, *an algorithm to obtain undistorted images* is executed [33]. This algorithm contains functions to compensate radial and tangential image deformity by using a predefined distortion model.

Subsequently, the *Kalman Filter* is applied to the persons detected to obtain their optimal pixel position. The Kalman Filter works in combination with the Munkres or Hungarian algorithm [35] for people track assignment. The latter algorithm is responsible for the association of trackers detected in a certain frame with the ones of the previous and subsequent ones based on pixel distance between frames. This procedure is extended in chapter 4.

Finally, the conversion from pixels to real coordinates in a common system for all the sensors considered is performed by using the *Perspective-n-Point* formula. As a result, the real 3D coordinates are computed as follows:

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (7)$$

where s is a scale parameter and u, v are the 2D projections. On the right side of Equation (7), the first matrix contains the intrinsic camera parameters: focal lengths f_x, f_y , and optical center c_x, c_y matching the 2D pixels with the 3D real world coordinates $[X, Y, Z, 1]'$. Furthermore, the second matrix, comprises the rotation r_{xx} and translation t_x values that change the coordinate reference to the desired system.

Kinect camera Tracking

Kinect camera incorporates three depth sensors and an algorithm for skeleton detection. This algorithm can be used to extract the person position in real 3D coordinates with respect to the camera focus [36].

Furthermore, the calibration process of this device can be performed by transforming the real coordinates $[X, Y, Z]'$ from Equation (7) to the desired common system reference. The process is summarized as follows:

- To get the calibration points an image is taken from Kinect camera to the floor markers.
- Using the open SDK from Kinect [22] the real world coordinates from camera coordinate system can be obtained by each selected point in the calibration.
- To move the coordinate system to the common coordinates, it is required to obtain the rotation matrix and translation vector:

Figure 8: Example of trajectories extraction from WSN (*left*), Kinect (*center*) and Zenith camera (*right*). The red line shows the real path whereas the green line represents the estimated path. The WSN estimation accuracy is lesser than the one obtained from cameras. However, WSN coverage area is the largest of the sensors chosen. Finally, the Kinect range of detection constraints the potential applications. In the experiment, WSN is composed of 4 nodes located at the top corners of the room. Kinect sensor is located at 500, 300 in horizontal and vertical axis respectively and Zenith Camera is located in the room center at ceiling.

Figure 9: Comparison of the estimated heading for bracelet and skeleton. The blue line is the raw data whereas the orange line represents the signal filtered. The skeleton noise is higher than the bracelet due to two main factors: (a) changes between 360 degrees and 0 are reflected in spite that variations are small; and (b) the short-term angle variation in the skeleton joints is large.

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = R \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + t \quad (8)$$

Both, the rotation matrix and the translation vector are obtained by applying a singular value decomposition (SVD) over the selected points.

$$U, S, V = \text{svd}((A - \bar{A})' * (B - \bar{B})) \quad (9)$$

$$R = V * U' \quad (10)$$

$$T = -R * \bar{A}' + \bar{B}' \quad (11)$$

where A is a $(n \times 3)$ matrix with n denoting the number of selected points and the column vector X, Y, Z contains the coordinates of the depth sensor. Moreover, B is a $(n \times 3)$ matrix with the x, y, z coordinates of the selected points in the reference coordinate system. $\bar{A}$ and $\bar{B}$ are the centroid values of x, y, z for all selected points.

3.3.4 Person Identification and Multi-modal Fusion

ABD is a fundamental part of the proposed system since this module is in charge of the data analysis to extract trajectories from data sensors and make inferences about elderly people actions. To reach such objective, a multi-modal scheme is proposed to fuse the information shown in Table 2.

However, in order to combine the information extracted from sensors, a previous *identification* stage must be performed. The reason is that multiple factors such as target occlusions, lighting changes, signal attenuation, reflections, among other effects, affect the tracking algorithms yielding to estimation errors. In fact, for scenarios such as the ones considered in this work, affluence of several persons is assumed and therefore a classification tool to associate the routes estimated with the corresponding person is required. Furthermore, the identification task must be performed for each sensor which implies that several matching levels from sensors data are needed.

In this work three main inference levels are proposed:

- *Track-to track comparison*: using the trajectories from Bracelets, Kinect camera and Zenith camera.
- *Orientation comparison*: based on heading estimation from bracelets and estimated orientation from Kinect and Zenith camera using the trajectories.
- *Accelerometer measures comparison* with wrist body point from Kinect camera.

Because of the system modularity, the identification process must be performed even if there is no data gathered from all sensors due to occlusions or coverage. Conversely, if measurements from Bracelets, Kinect and Zenith cameras are available, the three comparison methods can be applied to associate the estimated trajectories to a person in the scene.

Track-to-Track comparison

Assuming that data gathered have been synchronized in the previous stages of the low level subsystem, the goal of Track-to-Track procedure is to detect the routes with the larger similarity. The metric employed is the Minimum Square Error MSE along the route duration and can be expressed as follows:

$$\begin{aligned} &\text{minimize} \quad t \\ &\text{subject to} \quad f_i(\mathbf{x}) \leq t \quad i = 1, \dots, m. \end{aligned} \quad (12)$$

$$f_i(\mathbf{x}) = \frac{1}{M} \sum_{j=1}^M \sqrt{(x_j^i - x_j^j)^2 + (y_j^i - y_j^j)^2} \quad (13)$$

where x, y are the axis coordinates of a predefined route, m represents the number of routes in a scene (frame or bracelet time interval). M is the number of trajectory points taken into account in the estimation (corresponds to the time interval where synchronization is performed). The cost Matrix can be created with the MSE combinations between the mentioned sensors. As an example, if Kinect estimates two trajectories and Zenith camera three, the cost matrix is determined as:

$$C = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \end{bmatrix} \quad (14)$$

where every t_{ij} is obtained by solving the optimization problem (12). In this work, Munkres algorithm [35] is applied to the matrix (14). If all conditions are satisfied, the final assignment can be carried out and therefore link the trajectories for a determined person.

Orientation Comparison

Orientation comparison is proposed as second level of inference for person identification. Bracelets are equipped with sensors that directly provide the heading orientation. Moreover, there are several techniques to obtain the orientation based on the movements of the object tracked [49, 36].

The comparison will be performed using the metric described in (12). However, the level of comparison in this case is higher. The reason is that the orientation of the bracelets is not absolute when these devices lack of magnetometer. Therefore, a comparison of the orientation with the route rotated in the 360 degrees must be introduced in this process. The filtering process of the obtained curves is then executed to obtain the MSE. Analogously to the previous inference level, Hungarian algorithm [35] is utilized for the assignation process.

Acceleration comparison

The third inference level relies on the Bracelets accelerometer and wrist acceleration from Kinect body joints.

The process consists of extracting the data from every frame in Kinect raw. In concrete, wrist points (6,18 for left and right wrist respectively) from skeleton body must be compared to the acceleration values (ax, ay, az) extracted from the bracelets. Depending on the technology employed, the inertial measurements can include the gravity force (expressed in g). Therefore, initial normalization procedure must be performed before Kinect body and bracelets analysis. The association process can be then carried out as described in previous inference levels.

Multi-Modal Fusion

Once the identification procedure is performed, the measures are fused in a multi-modal scheme to improve the accuracy of estimations. A fusion algorithm can be used to achieve a high level of accuracy estimation using information from cameras, WSN and bracelets. In this work, Kalman Sensor Group Fusion algorithm (KSFG) is employed [28]. This algorithm exploits the Kalman filter properties for multi-sensor data fusion estimating a final x, y coordinates based on several trajectories. In particular, for the case of the sensor considered in this work, the filter has the following form:

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad (15)$$

Table 3: Data sources and outputs of the Clinical Decision Support System.

Data Source	Inputs	Output
Bracelets and Belts	Accelerometer Gyroscope BioSensors	Gait Evolution Heart Rate Temperature
ABD: Patient leaving home	WSN estimation	Warning generation.
ABD: Festination	evolution in bracelets and camera movements	Warning generations, number of Festination episodes over time.
ABD: Falls Detection	peaks in bracelets and kinect movements	Warning generations, number of falls over time.
ABD: Daily Motion	Movement evolution	Time spent performing daily activities, apathy.
Interfaces	Medical Annotations	Comorbidities, General Practitioner notes, medical history.
Interfaces	Elderly people interactions	i.e. Time surfing on interfaces, difficulties to use it, among others.
External sources	Information that supports medium-term decisions PaaS	Weather conditions can explain why patients stay at home, etc. Traffic information can be useful to schedule activities, etc.

$\mathbf{H}$ is the state matrix for Kalman filter. Furthermore, $\mathbf{R}$ matrix includes sensor error covariance measurements. These values are appended to the main diagonal:

$$\mathbf{R} = \begin{bmatrix} \sigma_{x_1}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{y_1}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{x_2}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{y_2}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{x_3}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{y_3}^2 \end{bmatrix} \quad (16)$$

Finally, the observations vector $\mathbf{z}$ collects the x and y coordinates estimated from each sensor:

$$\mathbf{z} = [\hat{x}_1, \hat{y}_1, \hat{x}_2, \hat{y}_2, \hat{x}_3, \hat{y}_3]' \quad (17)$$

As an output of the KSFG stage, a high accuracy daily motion estimation is obtained.

3.4 High Level Subsystem

3.4.1 Electronic Health Record (EHR)

The EHR is the collection of medical and behavioral data of patients. Information such as Health annotations, medications, frequency of ABDs generation, information from sensors (low level Subsys-

Figure 10: Details of high level subsystem modules and technologies integrated into a smart city infrastructure.

tem) among others is stored in the EHR. This data is used by the recommendation engines in CDS to provide inferences to the related stakeholders. As a result, General practitioners, Neurologists and Physicians will have a more complete health status of the patient for appropriate decision making about patients treatment.

There exist some extended standards for EHR (also known as Virtual Medical Record) [51, 52]. However, for the presented work, the Fast Healthcare Interoperability Resources (FHIR) standard is adopted. The reason to choose this standard is the interoperability between legacy health care systems (as the low level Subsystem), making the process of delivering the information to multiple stakeholders simple. The FHIR API [52] implements the operations for interaction with the medical information stored based on a HTTP RESTful. This implementation is fully compatible with the first-catalogue Generic Enable (GE) [53]. The main advantage of this GE is that it allows the interoperability of the modules developed with the Next Generation Services Interface (NGSI).

3.4.2 Clinical Decision Support (CDS)

There is a large list of proposed commercial and research CDS systems which are continuously evolving by applying complex algorithms for the recommendation processes. These systems have been classified into multiple groups:

- *Machine Learning*: These methods rely on known techniques such as Neural Networks, Supported Vector Machines. The goal is to train the algorithms to learn from the data.
- *Knowledge Based*: These systems contain knowledge modules that provide the rules for decision-making based on the expertise of professionals.

Figure 11: Example of fall detection module using fusion of sensors. In top graph, accelerometer variations for fall inference are outlined. In bottom, Kinect skeleton fall detection is drawn on the body.

Figure 12: Accelerometer and Gyroscope data for Freezing detection. On the (*right side*), the Freezing of Gait is appreciated whereas on the (*left side*), accelerometer data is noisy and therefore it can not be observed.

Figure 13: Evolution of patient location within home over a 24-hour period. The time the patient is outside is not monitored by the system.

- *Graphical Representation:* these methods allow the visualization of data to verify the data behavior and trends. The extraction of inferences is controlled by the data evaluator
- *Data Mining:* This approach consists in analyzing the data through linguistic techniques such as Natural Language Processing (NLP) [50].
- *Hybrid Approaches:* These techniques employ several of the aforementioned methods combined for decision-making.

The main advantage of the proposed system is that it allows the integration of multiple data sources to provide recommendations. In Table 3, data inputs are described, as well as some initial outputs are outlined. External sources such as smart city services (weather, traffic and environmental data) can be used to increase the system Knowledge (e.g. avoid places with large pollution, etc). The proposed platform to support these services is based on Cosmos ecosystem. This GE is intended to deploy means for analyzing both batch or/and stream data to provide insights on such a data revealing new information that was hidden. Batch data is stored in advance, and latency is not extremely important when processing it.

3.5 Experimental results

3.5.1 Methodology

The methodological aspects of experiments described in this work have been designed according to the project research book. To summarize, the experiments were performed in several sessions with a total of 18 patients. The control group has been composed of elderly people without cognitive problems (APM volunteers). Additionally, the experimental group was formed by patients that were classified into three groups according to they the illness stage as: initial, moderate or severe. Diverse aspects concerning gender balance as well as related diseases. This classification is of particular interest for the

Table 4: Algorithms Accuracy. Comparison of different positioning and tracking methods and the associated average error. Person identification success percentage based on the number of person presented in the monitored room.

WSN: Average error (centimeters) in fingerprinting technique using different methods					
	Euclidean distance	Mahalanobis distance	Weighted K-NN	Gaussian Kernel	Exponential Kernel
802.15.1 Bluetooth	150.57	142.58	140.32	128.15	137.67
802.15.4 XBee	110.71	104.56	98.36	81.73	87.5
WSN Average error (centimeters) using Tracking methods					
	Kalman Filter		Extended Kalman Filter	Particle Filter	
802.15.1 Bluetooth	120.02		108.33	85.11	
802.15.4 XBee	78.47		63.82	54.75	
Kinect Average error (centimeters) in skeleton Tracking					
		Direct Kinect Skeleton		Smoothed Skeleton Tobit Kalman Filter	
2D Calibration		36.45		29.67	
3D Calibration		43.91		34.34	
Zenith Average error (centimeters) in Tracking people					
Kalman Filter		Extended Kalman Filter (Only in combination with WSN)		Particle Filter (Only in combination with WSN)	
45.89		43.98		39.07	
Tracking Fusion Average error (centimeters)			Identification Accuracy		
WSN + Zenith	36.98		1 person	98%	
WSN + Kinect	32.71		2 persons	94%	
Kinect + Zenith	31.54		3 persons	90%	
WSN + Kinect + Zenith	30.11		+3 persons	≤89%	

high level subsystem. However, for the physical detection of abnormalities it is neglected. Furthermore, PD patients were informed on the procedure as well as the type of data that was going to be collected. However, participants were not notified neither hypothesis nor expected results to ensure unbiasedness

of their behaviors and consequent of the results.

3.5.2 Experiments and Results

The first experiment was intended to show the identification and individualization process. It is important to highlight that bracelet-result orientation is not absolute with regard to cardinal system, there exist a phase-change with regard to Kinect orientation which is absolute with camera focus as origin. It is drawn in Figure 9 where the vertical axis denotes orientation moves (angles) over time for both bracelets and Kinect sensor. After calibration process, the orientation changes must be similar in both cases.

The second experiment consists in abnormal events detection by the use of the bracelets as described in subsection 3.3. The analysis of accelerometer measures from bracelets provide information about strong changes in the movement. Gyroscope produces data about the velocity of turning (angular speed). Therefore, falls events can be detected by the combination of these two sensors as shown in Figure 11. In this figure, at top image the raw measurements can be observed. The left side figure is noisy, therefore a smoothing stage is applied using a Butterworth filter on the right side. Moreover, the bottom figure depicts how the skeleton is detected.

Furthermore, another abnormal events can be detected by the proper combination of inertial and visual sensors. In particular, in Freezing event detection, the goal is to find Freezing of Gait (FOG) [49] patterns in the time evolution of accelerometer measurements. In Figure 12, a time-lag corresponding to a FOG is shown. Notice that for this purpose, the signal from accelerometers must not be filtered.

In Figure 13, the location of the patient in the home is shown for a day. The living room area was the area covered by the whole set of sensors. The detection was mainly obtained through the WSN (via bracelet). The accuracy of the detection is high as the resolution is not so large (a room). Kinect and Zenith camera were deployed in the living room.

Finally, numerical results of the activities detected by the sensors separately are drawn in Table 4. Patients were recorded while walking for around 1 minute. Additionally, the improvements reached by the multi-modal fusion are detailed. From top to bottom:

- A comparison of diverse position estimation methods for WSN in a route across the room employed for the experiments using fingerprinting technique. These values show the error in centimeters of the estimated positions with respect to the real patient location. The employment of Kernel functions reaches better estimation results. The reason is that these functions attain a better wireless signal propagation modeling than linear functions.
- A comparison of tracking methods for position estimation is shown. It can be observed the significant error reduction by applying these techniques. The Kalman filter mitigates the impact of outliers, specially for WSN estimation, due to use of previous estimations reduces the error of noisy measurements.
- Estimation accuracy for two methods in Kinect devices is compared. There are two scenarios considered: (a) *2D calibration*: this method consists in use a fix plane (e.g. vertical axis $z = 0$) for the calibration process. (b) *3D calibration*: This method employs points in all coordinates. As it can be observed, the former method results are better as the main interest is in obtaining the patient position location in the room without the particular interest of the patient height.

- Tracking accuracy of several methods using the zenith camera (deployed at roof) is compared. The results obtained by using Particle Filter improves the results compared to the rest of mentioned methods. It is important to highlight that the error is similar to the one of Kinect sensor. However, for the sake of simplicity, results are shown only for areas covered by all sensors.
- Results of multi-modal process are drawn. On the left column, the accuracy improvement for diverse sensor configurations (ensuring modularity) fused is shown. Finally, the identification process results for cases (up to 4 individuals in the scene simultaneously) are described.

4 Chapter 4: Indoor Positioning.

Chapter Summary

This chapter extends one of the main concepts outlined in the previous chapter. In concrete, this chapter focuses in the indoor positioning and tracking of individuals using Wireless technologies. These technologies allow to obtain the Received Strength Signal Indicator (RSSI) to apply statistical techniques to estimate the person position. Additionally, the Kalman Filter is employed to track the persons. The contents of this chapter are directly associated to a Q1 Ranked Journal “*Adaptive Fingerprinting in Multi-Sensor Fusion for Accurate Indoor Tracking*”, in IEEE Sensors Journal, vol. 17, no. 15, pp. 4983-4998, 1 Aug.1, 2017 [38].

4.1 State-of-the-Art

Indoor positioning and tracking have attracted an extensive research effort because of their usefulness for a broad range of applications such as audience pattern generation [23], costumer analysis in retail [24], surveillance [57], business related activities [58], healthcare behavioral monitoring [59] among others [60, 61].

The aim of these research topics is to obtain knowledge of indoor target behaviors to allow other systems make proper decisions. The general approach involves the deployment of a set of sensors to detect and track the objects. A very common approach consists in the employment of cameras. However, there are some scenarios where these devices can not be used due to poor/inadequate lighting conditions or privacy issues. Moreover, GPS estimation is successfully applied for outdoor tracking, however it lacks of accuracy for indoor environments. Alternatively, several proposals face this problem by implementing wireless sensors in combination with processing algorithms that combine data gathered from multiple sources. Diverse methods such as Time-of-Arrival (ToA) [62], Time Difference of Arrival (TDoA) [63], Angle-of-Arrival (AoA) [64] and Received Strength Signal Indicator (RSSI) measurements [32] have been proposed. The techniques more often employed are ToA and RSSI. On the one hand, ToA techniques use timestamps lags from a sent/received packet. These methods allow high precision in the final estimation with relative low mathematical background. However, hardware cost limitations in addition to strict synchronization requirements increase the system complexity making this approach hard to implement. On the other hand, RSSI values can be measured by a large number of wireless devices. Nonetheless, RSSI modeling is generally tedious due to the harsh propagation conditions and processing techniques to achieve high accuracy estimations can be complicated [71].

Diverse proposals have been presented in the literature for indoor tracking [66, 34, 68, 69]. Some works employ theoretical propagation models to estimate the distance in combination with deterministic methods for tracking [34]. However, these propagation models neglect the obstacles, which means a high error rate in estimations.

A very extended approach is based on the employment of *Fingerprinting* which characterizes the scenario by splitting it into cells and performing spatial sampling [66, 67]. This technique can be applied in combination with both deterministic and probabilistic methods. The former methods depend heavily on the resolution of the fingerprint cells: the shorter the number of fingerprint cells is, the lower the estimation accuracy is. The latter methods consider a few sensors introducing a high mathematical

treatment to fuse the information gathered [68, 69].

Moreover, several works consider fusion from various technologies to extract common features aiming to obtain a high level inference. Additionally, recent works incorporate an adaptive update stage to improve the algorithmic accuracy based on the new information collected [70, 72]. Actually, this is a growing research topic as it allows to adapt fingerprint-based technology to changes in environment conditions.

Therefore, motivated by the heavy estimation requirements in terms of indoor tracking precision, in this work is proposed a Multi-Sensor Fusion based on Adaptive Fingerprint (MUFAF) Algorithm. This algorithm takes advantage of the availability to track objects that are equipped with several sensors creating fingerprinting (i.e a Smartphone). Afterwards, MUFAF performs an accurate statistical detection and tracking stage to obtain the object routes and finally, a phase an adaptive fingerprint update technique will adjust the algorithm to scenario variations.

Figure 14: RSSI behavior and variance along distance for different wireless sensors. WIFI variation is lesser than other technologies due to the short-term attenuation for considered distances. Conversely, XBee example is easy to fit into a valid model due to its decreasing behavior.

4.2 Problem statement

Complex indoor environments have been considered, where harsh signal propagation conditions present a challenge for appropriate characterization. For this aiming, it is assumed an indoor scenario with a set $\mathbf{v} = \{v_1, v_2, \dots, v_N\}$ of Access Points (AP) located at Cartesian coordinates $\mathbf{x}_j = (x_j, y_j) \in \mathbb{R}^2$. Furthermore, let l be the number of sensors that each node is equipped with. In this work, sensors from three wireless technologies are considered: IEEE802.11.g/WI-FI [120], IEEE802.15.1/Bluetooth v4.0 [119], IEEE802.15.4/xbee and IEEE802.15.4/CC2420 [21]. These sensors will obtain the Received Strength Signal Indicator (RSSI), denoted by $\mathbf{z}$, from the monitored target. Consequently, it is denoted by $\mathbf{z}^l(v_i(\mathbf{x}_j))$ the RSSI measurements from the sensor l gathered by node v_i at location $\mathbf{x}_j$.

The main problems addressed in this work are three: (a) the accurate localization estimation of the target by means of a combination of available RSSI measurements sensed by the nodes; (b) tracking of the detected object in an indoor environment and (c) the improvement in trajectories accuracy by means of multi-sensor fusion of the different sensors estimations. To tackle these challenges, several considerations are taken:

- $\mathbf{z}$ is assumed to be an independent and identically distributed (i.i.d.) Random Variable.
- Every node v_i is able to gather a set of RSSI measurements ($\mathbf{z}$) from the aforementioned technologies l . Notice that it is possible to obtain $\mathbf{z}^l \rightarrow \emptyset$
- T is defined as the measurement period. All data obtained in this time window will be assumed to be synchronized.
- Notation $\mathbf{z}^l(v_i(\mathbf{x}_j))$ describes the measures sensed by node v_i at a location $\mathbf{x}_j$ from fingerprint stage. However, for TPE process, notation $\mathbf{z}(v_i)$ denotes the set of RSSI measurements gathered by node v_i at the location that will be estimated.

4.2.1 Main contributions

The key contribution of this work is the proposal of a complete system for indoor localization and tracking that overcomes the difficult conditions of the scenarios in terms of propagation interference. Moreover, MUFAF is scalable and outperforms existing techniques therefore it can be considered as an adaptive system fusing multiple data sources. The main technical contributions of this section are:

- Several multi-sensor architecture patterns for tracking systems are compared. Some of the most precise methods for indoor tracking are described and assessed in this work. Moreover, as a result, MUFAF Algorithm for indoor localization using multi-sensor fusion from various technologies.
- An Adaptive Fingerprint Update (AFU) procedure is proposed to ensure that MUFAF will change according to the varying scenario conditions. This algorithm is based on Kriging interpolation that is a statistical technique that does not depend on the field mean. This feature guarantees AFU stage to be unbiased.
- A study on indoor localization and tracking is provided. As a result a complete open dataset is presented. This dataset contains all required fingerprint and route measurements that allow scientific community to test and compare their algorithms the ones presented in this work.

Figure 15: (Left) Fingerprint cells distribution used in this work (Right). RSSI Histogram examples for each sensor in a particular cell. It can be observed that WiFi arises narrow range of RSSI values for the cell considered whereas bluetooth histogram probability distribution is wide spread.

The remainder of this section is organized as follows: In next subsection 4.3, Target Positioning Estimation (TPE) is detailed. Subsection 4.4 describes the Target Tracking Procedure (TTP). In subsection 4.5, the comparison of fusion pattern architectures is defined. In subsection 4.6 adaptive fingerprinting technique using Kriging is presented. Moreover, in subsection 4.7 the implementation details and results of the algorithms are illustrated.

4.3 Position Estimation

In this section, the algorithm used to obtain the object position in an indoor environment is described. This process comprises two steps: (a) Fingerprinting and (b) Kernel Density Estimation (KDE). The former is a method for scenario characterization while the latter is a statistical technique for position estimation applied to sensor RSSI measures. This algorithm employs the Fingerprinting data to estimate the target location.

Algorithm 1 MUFAF: Fingerprint Stage

```

1:  $\mathbf{v} = \{v_1, \dots, v_N\}$ : set of Access Point locations  $\mathbf{c} = \{c_1, \dots, c_\ell\}$ : set of fingerprint cells.
    $z(v_j(\mathbf{x}_k))$ : RSSI measured by Access Point  $j$  at location  $\mathbf{x}_k$ .  $\tau$  is the predefined fingerprinting
   period and  $t$  is a timer.
2: procedure FINGERPRINT TRAINING
3:   for each point  $i$  in  $\mathbf{v}$  do
4:      $\mathbf{x}_i = (x_i, y_i)$ 
5:     for each cell  $j$  in  $\mathbf{c}$  do
6:       for each technology  $l$  do
7:         while  $t - \tau > 0$  do
8:            $\mathbf{F}^l(v_i(\mathbf{x}_j)) \leftarrow z^l(v_i(\mathbf{x}_j))$ 
9:         end while
10:      end for
11:    end for
12:  end for
13: end procedure

```

4.3.1 Fingerprinting

Large variance of measurements gathered is one of the most critical problems associated to RSSI estimation, specially for indoor environments, even in case of static objects [72]. An example of this variance is shown in figure 14 where a large amount of RSSI measurements are taken from a static object equipped with several sensors. A common approach to deal with this issue is by applying Fingerprinting Technique [55, 56, 25]. This technique allows to obtain a proper scenario calibration by splitting it into a set of cells $\mathbf{c} = \{c_1, c_2, \dots, c_\ell\}$ with geometrical center $\mathbf{x}_j = (x_j, y_j) \in \mathbb{R}^2$ respectively. An example of this distribution is drawn in left side of figure 15. The larger the number of cells ℓ is, the higher the spatial resolution is. Measures are taken from each node v_i in the center of every cell $\mathbf{x}_j$ for a predefined time interval τ . The entire Fingerprint process is detailed in algorithm 1.

As a result of this procedure, a distribution of the RSSI is captured in every cell k for every technology l and every node v_i . For this purpose, it is defined $\mathbf{F}$ matrix that contains the fingerprinting distributions. Therefore, $\mathbf{F}^l(v_i(\mathbf{x}_j))$ represents the set of fingerprint measurements gathered for cell $\mathbf{x}_j \in \mathbb{R}^2 \wedge c_j \in \mathbf{c}$ by sensor technology l in node $v_i \in \mathbf{v}$. Moreover, $\mathbf{F}$ matrix is the input of the TPE process described in next subsection. An example of fingerprinting cells and RSSI fingerprinting distributions is shown in right side of figure 15.

4.3.2 Target Position Estimation (TPE)

There exist several works in the literature for position estimation based on Fingerprint statistics [83, 86, 87, 88]. Most of these works apply deterministic techniques to characterize the RSSI distribution for each cell. In [83], the K-Nearest Neighbours (KNN) technique is employed to triangulate the target position adopting euclidean distance as metric. In [86], a comparison of various distance metrics is carried out in an indoor building giving as result that Mahalanobis distance outperforms Manhattan and Euclidean-Distance results. Furthermore, in [87] a weighted K-Nearest Neighbour approach is presented. Moreover, in [88] it is proposed a complete overview of several deterministic methods that include all the aforementioned metrics. However, taking into consideration a large set of measures

collected for every cell in Fingerprint stage, a deterministic approach can neglect the wide distribution of measurements per cell, therefore a probabilistic approach can be considered for the estimation. In particular, Bayes theorem based methods can be applied for position estimation as follows:

$$p(\mathbf{x}|\mathbf{z}) = \frac{p(\mathbf{x}, \mathbf{z})}{p(\mathbf{z})} \quad (18)$$

where $p(\mathbf{x}|\mathbf{z})$ is the conditional probability and $p(\mathbf{x}, \mathbf{z})$ is the joint probability.

Kernel Density Estimation (KDE) is a non-parametric technique employed to obtain the Probability Density Function (PDF from now on) of a random variable with independent and identically distributed samples [58, 59]. KDE estimates the likelihood that Fingerprinting distributions match measures in the region of a Kernel function. Therefore, the distance is fitted with Kernel functions, denoted by $K(\cdot)$, that must fulfill the following conditions:

$$K(-\mathbf{x}) = K(\mathbf{x}) \quad (19)$$

$$\int K(\mathbf{x})d\mathbf{x} = 1 \quad (20)$$

equation (19) states that Kernel functions must be non-negative. Moreover, equation (20) implies that Kernel functions, as well as the PDF must be unbiased (therefore be integrable to one). If both conditions are satisfied, the PDF can be calculated in its general form by the following function:

$$p(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \frac{1}{\mathbb{V}} K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h_n}\right) \quad (21)$$

where n is the set of samples. Furthermore, $\mathbb{V}$ is a normalization parameter of the Kernel volume and h_n is a smoothing parameter also known as kernel bandwidth. The goal pursued is to estimate a non-parametric function that fits the RSSI values distribution better than known (parametric) distributions. Specifically, equation (21) can be applied to our fingerprinting by employing histograms to smooth the measurements distribution. Therefore, the maximum likelihood function can be expressed as:

$$p(\mathbf{z}(v_i)|c_j) = \frac{1}{nh} \sum_{k=1}^n K\left(\frac{\mathbf{z}(v_i) - \mathbf{z}_k(v_i(\mathbf{x}_j))}{h_n}\right) \quad (22)$$

where n is the number of measurements gathered in every fingerprinting cell $j \in \mathbf{c}$ by the node $v_i \in \mathbf{v}$, $\mathbf{z}(v_i)$ is the set of RSSI measures gathered by node v_i in a window time T . In addition, $\mathbf{z}(v_i(\mathbf{x}_j))$ is the set of RSSI values gathered by node v_i in fingerprint cell located at x_j . Due to the nature of the data modeled, a non-parametric method is chosen to calculate the joint density probability function. In particular, a modified version of KDE known as Nadaraya-Watson Kernel regression is used [28]. This method is applied because of its appropriateness when there is no prior knowledge of the relationship between the variables under study since these estimators are only based on either smoothing or regression functions.

There are other non-parametric options that could have been employed such as the Priestley-Chao and Gasser-Müller smoothers or the K-th Nearest-Neighbor (K-NN) weights [84]. However, Priestly-Chao and Gasser-Müller strict boundary bias problems make Nadaraya-Watson method the most suitable estimator for RSSI-based applications. In fact, it has been demonstrated that Nadaraya-Watson variance is up to a 50% lesser than the other methods [84]. However, K-NN smoothing can attain similar results to the ones obtained by using kernel estimation when an appropriate bandwidth h parameter is adjusted.

According to this method, the joint probability function can be estimated as:

$$p(\mathbf{x}, \mathbf{z}(v_i)) = \frac{1}{\ell} \sum_{k=1}^{\ell} \frac{1}{(h_{\mathbf{x}})^2 (h_n)^N} K\left(\frac{d(\mathbf{x}_k, \mathbf{x})}{h_{\mathbf{x}}}\right) K\left(\frac{\mathbf{z}(v_i) - \mathbb{E}[\mathbf{z}(v_i(\mathbf{x}_k))]}{h_n}\right) \quad (23)$$

where $d(\mathbf{x}_k, \mathbf{x})$ represents the distance from every fingerprinting cell center to the estimated point and $h_{\mathbf{x}}, h_n$ are the bandwidth parameters respectively.

The most common Kernel functions used to fit the PDFs in equation (24) are the Gaussian and Exponential:

$$K(\mathbf{x}) = (2\pi)^{-\frac{N}{2}} |\mathbf{C}|^{-\frac{1}{2}} \exp(-\frac{1}{2} \mathbf{x}' \mathbf{C}^{-1} \mathbf{x}) \quad (24)$$

$$K(\mathbf{x}) = \frac{1}{2} \exp(-\|\mathbf{x}\|) \quad (25)$$

The former is the Gaussian Kernel (24), where N is the number of nodes considered for estimation, whereas $\mathbf{C}$ represents the covariance matrix which is a diagonal matrix with the covariances of the N node measurements. The equation (25) is the Exponential Kernel where $\|\mathbf{x}\|$ represents the euclidean distance.

Applying the Minimum Mean Square Error (MMSE) criteria over the conditional density probability:

$$\hat{\mathbf{x}}_{MMSE} = E(\mathbf{x}|\mathbf{z}(v_i(\mathbf{x}))) = \int_C \mathbf{x} p(\mathbf{x}|\mathbf{z}(v_i(\mathbf{x}))) dx \quad (26)$$

it is obtained that:

$$\hat{\mathbf{x}} = \sum_{k=1}^{\ell} w_k \mathbf{x}_k \quad (27)$$

where $\mathbf{x}_k \in \mathbb{R}^2 \wedge k \in \mathbf{c}$. The target position $\hat{\mathbf{x}}$ is calculated as the weighted sum of the estimations for

every cell using the Kernel function. The weights w are determined by:

$$w_k = \frac{K\left(\frac{\mathbf{z}(v_i) - \mathbb{E}[\mathbf{z}_k(v_i(\mathbf{x}_k))]}{h_n}\right)}{\sum_{j=1}^{\ell} K\left(\frac{\mathbf{z}(v_i) - \mathbb{E}[\mathbf{z}_j(v_i(\mathbf{x}_j))]}{h_n}\right)} \quad (28)$$

The numerator contains the current Kernel for cell $k \in \mathbf{c}$ and the denominator collects the sum of all Kernel values for every cell. The Nadaraya-Watson Kernel regression is the MMSE estimator of $\mathbf{x}$.

Finally, the TPE output is the estimated value $\hat{\mathbf{x}}$ which is assumed to be static in T . Moreover, in next subsection, the tracking algorithm to obtain the complete route is drawn.

Figure 16: Kalman Filter Algorithm steps for prediction and estimation.

4.4 Target Tracking Process (TTP)

Tracking stage is intended to minimize the effect of noise in $\hat{\mathbf{x}}$ by filtering outliers giving as a result the peaks-free object route. Several methods have been proposed for this purpose, however in this work the well-known Kalman Filter (KF) [73, 74] is employed. This filter provides a recursive solution based on Least Squares Method.

KF assumes the system to be described by a linear stochastic model, where the error associated to the system as well as the additional information incorporated are presented as normal distribution variables with zero mean and variance σ^2 .

A complete illustration of Kalman Filter is drawn in figure (16). This filter performs the Best Linear Unbiased Estimator (B.L.U.E) of a system state using information from the previous estimations. KF

Figure 17: Architecture patterns for sensor fusion. (Up). *Track-to-track Fusion*: Separate processing stages for each sensor to achieve high-level inferences that are subsequently fused. (Bottom) *Kalman Sensor Group Fusion*: direct sensor data combination.

consists of two stages: (a) time update (left) and (b) measurement update (right) are related. The former is in charge of projecting future estimators of the current state and error covariance. The latter will predict new state estimations.

One of the most interesting KF features is the ability to predict the state of a system, even if the nature of the system modeled is unknown. The *Time Update (prediction)* estimates the state $\mathbf{x}$ according to the following equation:

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}_k u_k + \eta_k \quad (29)$$

where $\eta \sim \mathcal{N}(0, \sigma)$ is the process noise, $\mathbf{A}$ matrix relates the k state with the $k + 1$ time state. Furthermore, $\mathbf{B}$ is the control input associated to u , which is the system input. Furthermore, $\mathbf{P}$ collects the deviation in the estimates that generally is initialized as a diagonal matrix and $\mathbf{Q}$ is the covariance matrix.

Moreover, *Measurement update (prediction)* stage in figure 16 is described as follows:

- (1): The matrix $\mathbf{G}$, called Kalman Gain or blend factor, adjusts the trade-off between the estimation and the error measurement.
- (2): Equation to calculate the posteriori state $\mathbf{x}_{k+1}$ as a linear combination of the a priori $\mathbf{x}_k$ estimator and the weighted difference between current observations $\mathbf{z}_k$ and a measurement prediction $\mathbf{H}$.
- (3): Update the error covariance based on the Kalman Gain. Moreover, the observation $\mathbf{z} \in \mathbb{R}^2$ contains the spatial coordinates of the position :

$$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + v_k \quad (30)$$

Finally, matrix $\mathbf{H}$ relates the state with the measure $\mathbf{z}_k$ where $\mathbf{z}_k$ is the TPE output and $v \sim \mathcal{N}(0, \sigma)$ the measurement noise respectively.

In this work, a Modified version of the Kalman Filter (MKF) [75] is applied. The reason to use MKF is that the model considered in this system is linear. Consequently, a linear version of KF taking into consideration velocity and acceleration parameters represents a good trade-off between complexity (i.e Extended Kalman Filter) and performance (velocity and acceleration improve the estimation results). In addition, Extended Kalman Filter does not provide the optimal solution in case of wrong initial state inputs, which is highly probable in the scenarios considered in this work [89].

Therefore, MKF algorithm considers the speed axis v_x and v_y variables for the state $\mathbf{x}$ in addition to the position vector $\hat{\mathbf{x}} \rightarrow (\hat{x}, \hat{y})$ of the tracked target. These variables represent an advantage in terms of estimation accuracy since speed endows the system with a more realistic movement modelling. Therefore, the system can be expressed as follows:

$$\mathbf{x} = [x, y, v_x, v_y]' \quad (31)$$

$\mathbf{x}$ is the initial state composed by axis positions and speed. Consequently, matrix $\mathbf{A}$ will be defined as:

$$\begin{cases} x_{k+1} = x_k + v_x dt \\ y_{k+1} = y_k + v_y dt \\ v_{x_{k+1}} = v_{x_k} \\ v_{y_{k+1}} = v_{y_k} \end{cases} \quad \mathbf{A} = \begin{bmatrix} 1 & 0 & dt & 0 \\ 0 & 1 & 0 & dt \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (32)$$

Additionally, $\mathbf{H}$ remains as the general Kalman formulation whereas $\mathbf{B}$ and u_k will be modified as follows:

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \frac{1}{2} dt^2 \\ \frac{1}{2} dt^2 \\ dt \\ dt \end{bmatrix} \sigma_a^2 \quad (33)$$

Main modifications presenting the movement information and are included in different matrices of the

process, where $\sigma_a^2 = u_k$ is the acceleration parameter. Moreover, $\mathbf{R}$ matrix contains the error variance of TPE estimation:

$$\mathbf{R} = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix} \quad (34)$$

where σ_x^2, σ_y^2 represent the error variance estimation in x and y respectively. Finally, covariance matrix $\mathbf{Q}$ represents the uncertainty in the estimation process incorporating the entire set of parameters and variance σ_q^2 :

$$\mathbf{Q} = \begin{bmatrix} \frac{1}{4}dt^4 & 0 & \frac{1}{2}dt^3 & 0 \\ 0 & \frac{1}{4}dt^4 & 0 & \frac{1}{2}dt^3 \\ \frac{1}{2}dt^3 & 0 & dt^2 & 0 \\ 0 & \frac{1}{2}dt^3 & 0 & dt^2 \end{bmatrix} \sigma_q^2 \quad (35)$$

4.5 Multi-Sensor Processing Architectures

In this work, mainly two multi-sensor schemes are considered: (1) Track-to-track fusion and (2) Kalman Sensor Group Architecture. Both architecture patterns are shown in figure 17. In the former, processing stages (TPE and TTP) are performed separately for each sensor to achieve high-level inferences (routes) that are subsequently fused. Nonetheless, in the latter scheme the sensor measurements are combined in the TTP stage. Both schemes are described and compared in next subsections and results are presented in section 4.7.

4.5.1 Track-To-Track Fusion Architecture TTTF

This architecture pattern is based on fusing the routes obtained by each sensor l separately. The algorithm proposed applies the Maximum Likelihood Estimator (MLE) to the probability density function of the available routes [76]. For this purpose, $\mathfrak{R}$ is defined to be the set of routes. Additionally, it is assumed that every route is composed by ϖ number of steps $\mathfrak{R}^l = \{\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_\varpi\}$. Therefore, the probability density function can be described as follows:

$$p(\mathfrak{R}^l) = \prod_{k=1}^{\varpi} \prod_{n \in l_k} \frac{1}{\sqrt{(2\pi)^d |\mathbf{P}_n|}} e^{-\frac{1}{2}(\hat{\mathbf{x}}_n - \bar{\mathbf{x}}_k)' \mathbf{P}_n^{-1} (\hat{\mathbf{x}}_n - \bar{\mathbf{x}}_k)} \quad (36)$$

where $\mathbf{P}$ and $\mathbf{x}$ are the outputs from *TTP* described in section 4.4. Note that $\hat{\mathbf{x}}$ describes the position estimated from previous stages, whereas $\bar{\mathbf{x}}$ denotes the final output of the Fusion stage. From equation (36), it has been demonstrated [90] that MLE can be obtained by minimizing the result of $\mathbf{x}$ and

setting the gradient equal to 0. Therefore, the resulting log-likelihood function is:

$$\begin{aligned} \Lambda(\varpi, \{\mathfrak{R}^l\}, \{\mathbf{x}_k\}_{k=1}^{\kappa}) = & \\ & \frac{1}{2} \sum_{k=1}^{\varpi} \sum_{n \in l_k} \left(\ln((2\pi)^{\mathfrak{d}} |\mathbf{P}_n|) \right. \\ & \left. + (\hat{\mathbf{x}}_n - \bar{\mathbf{x}}_k)' \mathbf{P}_n^{-1} (\hat{\mathbf{x}}_n - \bar{\mathbf{x}}_k) \right) \end{aligned} \quad (37)$$

where $\mathfrak{d}$ is the dimension. Moreover, the state estimation for every location can be calculated as follows:

$$\bar{\mathbf{x}}_k = \left(\sum_{n \in l_k} \mathbf{P}_n^{-1} \right)^{-1} \sum_{n \in l_k} \mathbf{P}_n^{-1} \hat{\mathbf{x}}_n \quad (38)$$

Note that, both the estimation $\bar{\mathbf{x}}_k$ and $\mathbf{E}(\bar{\mathbf{x}}_k) = \bar{\mathbf{x}}_k$ are unbiased. Finally, the covariance of the estimate, denoted as Σ , is given by:

$$\Sigma_{\varpi} = \left(\sum_{n \in l_k} \mathbf{P}_n^{-1} \right)^{-1} \quad (39)$$

It can be shown that this result is the Kalman Filter update equation applied to each local route in each partition. Further details can be found in [90].

4.5.2 Kalman Sensor Group Fusion Architecture KSGF

The second architecture presented is the Kalman Sensor Group Fusion (KSGF) which exploits the Kalman filter properties for multi-sensor data fusion [77]. KSGF method directly incorporates the position estimations of each sensor in a single Kalman Filter. However, the system must be properly conditioned to obtain an accurate route estimation.

General KF system equations were described in section 4.4. However, some modifications must be introduced in the filter to support multi-sensor functionalities. In concrete, update matrix $\mathbf{H}$, described in equation (33), must be adapted to incorporate the sensor inputs:

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad (40)$$

Algorithm 2 MUFAF: TPE, TTP, TTTF Stages

```

1: for each measure  $\mathbf{z}(v_i)$  gathered and previous estimation  $\hat{\mathbf{x}}_{k-1}$  in  $\tau$  do
   Target Position Estimation (TPE) variables:  $\mathbf{v} = \{v_1, \dots, v_N\}$ : set of Access Point loca-
   tions /  $\mathbf{c} = \{c_1, \dots, c_\ell\}$ : set of fingerprint cells. /  $w_i$ : weight assigned to point  $i$ . /  $K(\cdot)$ : Ker-
   nel function. / Inputs:  $\mathbf{x}_k : (x_k, y_k)$  center of fingerprint cells /  $\mathbf{z}(v_i)$ : RSSI measurement. /
    $\mathbb{E}[\mathbf{z}_j(v_i(\mathbf{x}_j))]$ : Expectation of measurements in a fingerprint cell / Outputs:  $\hat{\mathbf{x}}_j : (\hat{x}_j, \hat{y}_j)$  location
   estimated /
2:   procedure TPE:
3:      $p_j \leftarrow \sum_{j=1}^{\ell} K\left(\frac{\mathbf{z}(v_i) - \mathbb{E}[\mathbf{z}_j(v_i(\mathbf{x}_j))]}{h_n}\right)$ 
4:      $\hat{\mathbf{x}}_j \leftarrow \sum_{k=1}^{\ell} \frac{K\left(\frac{\mathbf{z}(v_i) - \mathbb{E}[\mathbf{z}_j(v_i(\mathbf{x}_k))]}{h_n}\right)}{p_j} \mathbf{x}_k$  ; (27)
5:   end procedure
   Target Tracking Variables:  $\mathbf{P}$  Error Covariance Estimation Matrix /  $\mathfrak{R}^l$ : Set of routes per tech-
   nology  $l$  / Inputs:  $\hat{\mathbf{x}}_j$ : position estimated for each technology  $l$  in TPE.
6:   procedure TTP:
7:     for each  $\mathfrak{R}^l$  do
8:        $\hat{\mathbf{x}}_{k+1}^-, \hat{\mathbf{P}}_{k+1}^- \rightarrow$  Kalman Time Update Fig 16
9:        $\hat{\mathbf{x}}_k, \hat{\mathbf{P}}_k \rightarrow$  Kalman Measurement Update Fig 16
10:    end for
11:  end procedure
   Track-To-Track Fusion:  $\mathfrak{R}^l$ : Set of routes per technology  $l$  / Inputs:  $\mathbf{P}$ : Error Covariance Es-
   timation Matrix /  $\hat{\mathbf{x}}_k$ : position estimated for each technology  $l$  in TPE. Outputs:  $\bar{\mathbf{x}}_k$ : position
   estimated in Fusion Stage
12:  procedure TTTF:
13:    for each  $\mathfrak{R}^l$  do
14:       $\bar{\mathbf{x}}_k \leftarrow \left(\sum_{n \in l_k} \mathbf{P}_n^{-1}\right)^{-1} \sum_{n \in l_k} \mathbf{P}_n^{-1} \hat{\mathbf{x}}_k$ 
15:    end for
    $\alpha - \beta$  filter:
16:    Calculate predictions for position: (43)
17:    Compute prediction error: (44)
18:    Estimate the position: (45)
19:    Estimate the velocity: (46)
20:  end procedure
21:   $\hat{\mathbf{x}}_{k-1} \leftarrow \bar{\mathbf{x}}_k$ 
22: end for

```

Furthermore, $\mathbf{R}$ matrix must include sensor error covariance measurements. These values are appended to the main diagonal:

$$\mathbf{R} = \begin{bmatrix} \sigma_{x_1}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{y_1}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{x_2}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{y_2}^2 & 0 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{x_l}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{y_l}^2 \end{bmatrix} \quad (41)$$

Finally, the observations vector $\mathbf{z}$ is now extended to collect the x and y coordinates estimated from each sensor:

$$\mathbf{z} = [\hat{x}_1, \hat{y}_1, \hat{x}_2, \hat{y}_2, \dots, \hat{x}_l, \hat{y}_l]' \quad (42)$$

Taking into consideration the aforementioned modifications, the KF can be performed as described in *TTP*. As a result, the fused route will be provided by the KSGF. Both architecture patterns can be employed for multi-sensor data fusion. The computational complexity of the architectures proposed has been addressed [78]. In this work, both patterns are compared in terms of execution time. KSGF computing time grows exponentially as the number of sensors increases. In addition, TTTF scheme outperforms KSGF in terms of estimation accuracy. Therefore, in this work TTTF architecture is adopted for estimation process.

4.5.3 Alpha-Beta Filter for noise reduction

In order to minimize the noise and outliers effect that can arise as result of multi-sensor stage, an additional filtering stage is applied. There exist several proposals for routes smoothing, however in this work, we use the well known alpha-beta filter [79]. The reasons to employ the alpha-beta filter are the lower complexity, the lesser computational cost and that system model details are not required. Furthermore, taking into consideration that the effect of outliers is much lesser in the presence of low noisy estimations, it has been shown that the alpha-beta filter estimations performance is similar to Kalman Filters [92].

This filter is a particular case of the KF described in previous sections and consequently it assumes that a system can be adequately approximated by a two states model [79, 80].

Similarly to equation (31), initial states of spatial coordinates and axis speeds are considered. x, y, v_x, v_y . Therefore, the position prediction can be obtained as follows:

$$\hat{x}_{k+1} = \hat{x}_k + \hat{v}_{x_k} t ; \hat{y}_{k+1} = \hat{y}_k + \hat{v}_{y_k} t \quad (43)$$

Additionally, error measurements can be calculated from predicted states by the subtraction of the

fusion algorithm $\hat{x}_m, \hat{y}_m$ from the predicted position:

$$e_x = \hat{x}_m - \hat{x}_{k+1} ; e_y = \hat{y}_m - \hat{y}_{k+1} \quad (44)$$

The final position and velocity for next algorithm iterations are obtained as follows:

$$\hat{x}_k = \hat{x}_{k+1} + \alpha e_x ; \hat{y}_k = \hat{y}_{k+1} + \alpha e_y \quad (45)$$

$$\hat{v}_{x_k} = \hat{v}_{x_k} + \frac{\beta}{\Delta t} e_x ; \hat{v}_{y_k} = \hat{v}_{y_k} + \frac{\beta}{\Delta t} e_y \quad (46)$$

where α and β are the filter control parameters. In order to guarantee filter stability and convergence, these values must fulfill the following conditions:

$$0 < \alpha < 1; 0 < \beta \leq 2; 0 < 4 - 2\alpha - \beta \quad (47)$$

Additionally, there is a direct relation between alpha-beta control parameters:

$$\beta = \frac{\alpha^2}{2 - \alpha} \quad (48)$$

Figure 18: Main Semivariogram Patterns: *Nugget effect*: uncertainty due to lack of very close samples. *Range*: larger distances than Range are uncorrelated. *Sill*: maximum similarity value.

The entire procedure for MUFAF Multi-Sensor Fusion is described in algorithm 2. In this pseudo-code the sequence to perform the TPE, TTP and TTTF stages are shown for an iteration.

Algorithm 3 MUFAF: AFU

Adaptive Fingerprinting Update $\hat{\mathbf{z}}(\cdot)$: RSSI estimation. / $\bar{\gamma}(\cdot)$: SemiVariogram function, / $\mathbf{\Gamma}$: Semivariogram Matrix. /

- 1: **procedure** AFU:
- 2: **for each** point i, j in $\mathbf{c}$ **do**
- 3: Calculate $\hat{\gamma}(d(\mathbf{x}_i, \mathbf{x}_j))$; (49)
- 4: **end for**
- 5: Fit into theoretical model $\bar{\gamma}(\cdot)$
- 6: Create $\mathbf{\Gamma}$ Matrix
- 7: Create $\mathbf{b}$, $\mathbf{b}_{c+1} \leftarrow 1$
- 8: Solve SLE $\mathbf{w} \leftarrow \mathbf{\Gamma}^{-1}\mathbf{b}$; (54)
- 9: Obtain Estimation $\mathbf{z}(\hat{\mathbf{x}}_j) \leftarrow \sum_{l=1}^M \lambda_l \mathbf{z}(\mathbf{x}_c)$; (53)
- 10: Apply Rule: $\frac{1}{N} \sum_{i=1}^N \|\hat{z}(v_i(\hat{\mathbf{x}})) - z(v_i(\hat{\mathbf{x}}))\| \leq \mu$
- 11: **end procedure**

4.6 Adaptive Fingerprinting Update

In this section, the Adaptive Fingerprint Update (AFU) stage that endows MUFAF with responsiveness to indoor condition changes by adjusting the Fingerprint collection is detailed. For this aiming, an interpolation technique called Kriging [80, 81] is employed. This technique is mainly used because Kriging interpolation is performed even if the mean of the field is unknown which represents an advantage over traditional methods based on covariance. Therefore, exploitation of spatial similarity is performed through a statistical tool called the semivariance, which is more robust to outliers.

This technique was initially proposed for geospatial analysis, however the method has been successfully applied to several sciences. The main goal of Kriging is to estimate the field behaviour at unknown locations based on the available measurements.

The semivariance is a relevant tool used in Kriging interpolation, and its general estimator is given by the experimental Variogram, denoted by $\hat{\gamma}(\cdot)$:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i,j=1}^{N(h)} (\mathbf{z}(x_i) - \mathbf{z}(x_j))^2 \quad (49)$$

where $N(h)$ represents the number of RSSI measurements gathered at a predefined distance h . Specifically, in MUFAF all fingerprint cells $\mathbf{c}$ are employed to build the semivariogram:

$$2\hat{\gamma}^l(\mathbf{h}) = \frac{1}{\|\mathbf{c}(\mathbf{h})\|} \sum_{\mathbf{c}(\mathbf{h})} (\mathbf{z}^l(\mathbf{x}_i) - \mathbf{z}^l(\mathbf{x}_j))^2 \quad (50)$$

$\forall i, j \in \mathbf{c} \wedge d(\mathbf{x}_i, \mathbf{x}_j) \in \mathbf{h}$. This process is performed for each AP ($\mathbf{z}^l(v_k(\mathbf{x}_i)) \forall v_k \in \mathbf{v}$). In the case of

Figure 19: Top graphics represent the initial histogram representation for a cell of 50 measurements whereas bottom shows the evolution of Histograms in 300 simulations provided by AFU (*Left*) Evolution of Histogram for AFU in Bluetooth. (*Center*). Evolution of Histogram for AFU in WiFi. (*Right*) Evolution of Histogram for AFU in 802.15.4.

square/rectangular grids as the ones considered in figure 18, the lags distribution can be expressed as:

$$d(x_i, y_j) = \sqrt{\left(\frac{d1}{2}c_{xi} - \frac{d2}{2}c_{xj}\right)^2 + \left(\frac{d1}{2}c_{yi} - \frac{d2}{2}c_{yj}\right)^2} \quad (51)$$

where $i, j = \{1, \dots, \ell\}$, $x_i, y_j \in \mathbf{c}$ are the spatial coordinates of the cells center. Furthermore, W and D are the indoor width and depth expressed in centimeters respectively. Additionally, $\ell = n \times m$ is the total number of cells and n and m are the number of cells per side (width and depth). Finally, $d1, d2$ denote the distance from the center of a cell to the center of a neighbor cell in every axis. Moreover, it has been shown [91] that if the number of intervals within lags in $\mathbf{h}$ satisfies to be lesser than $\sqrt[2]{\frac{\pi\sqrt{W^2+D^2}}{\log(\sqrt{W^2+D^2})}}$ then it will exist at least a semivariance measurement in every $\mathbf{h}$ interval.

Once the Experimental Semi-Variogram is calculated, a fitting into a general theoretical model, denoted by $\bar{\gamma}(\cdot)$ is required to infer spatial data for distances not considered in $\hat{\gamma}(\cdot)$.

There exist several theoretical models for the Semi-Variogram such as Linear, Spherical, Gaussian and Exponential [82]. Gaussian model is appropriate for process with small short-term variations of similarity whereas Spherical or Exponential are the most suitable options for applications where considerable variations arise in short-distance. Due to the high variance in measurements distribution

presented in section 4.2, in this work, Spherical Semivariogram theoretical model is employed:

$$\bar{\gamma}(h) = ng + Rg \left(\frac{3h}{2Si} - \frac{h^3}{2Si^3} \right) \quad (52)$$

where ng is called the *nugget effect*. This parameter considers the uncertainty of the Semivariogram estimation when the separation between points tends to 0. Rg is the *Range*, that is the maximum distance taken into consideration for two samples. Finally, the *Sill* (Si) is the maximum variance (dissimilarity) between two samples. These parameters are drawn in figure 18.

Kriging provides the Best Linear Unbiased Estimator (B.L.U.E). Therefore, the field value $\hat{\mathbf{z}}(\hat{\mathbf{x}})$ can be described as a weighted sum of the available node measurements:

$$\hat{\mathbf{z}}(\hat{\mathbf{x}}) = \sum_{i=1}^N \lambda_i \hat{\mathbf{z}}(\mathbf{x}_i) \quad (53)$$

where $\lambda = \{\lambda_1, \dots, \lambda_N\}$ are the weights which sum must fulfil $\sum \lambda = 1$. Moreover, the weights are obtained by solving the following System of Linear Equations (SLE):

$$\mathbf{\Gamma} \mathbf{s} = \mathbf{b} \quad (54)$$

where:

$$\mathbf{\Gamma} \equiv \begin{cases} \bar{\gamma}(d(\mathbf{x}_i, \mathbf{x}_j)), & i, j = 1, \dots, N \\ 1, & i = N+1, j = 1, \dots, N; \\ & j = N+1, i = 1, \dots, N \\ 0, & i, j = N+1 \end{cases}$$

Figure 20: (*left*) Path estimation by using Track-To-Track Fusion Architecture and (*right*) Path estimation by using Kalman Sensor Group Fusion Architecture.

Figure 21: Average error comparison of route estimation for every technology (no fusion applied) using Fingerprinting technique, Kernel Density Estimation and Kalman Filter.

is an $N + 1, N + 1$ Matrix capturing the spatial similarity among the field. Additionally:

$$\mathbf{s} = [\lambda, \mathcal{L}]'$$

$\mathbf{s}$ vector contains the weights λ and the $N + 1$ value is the Lagrange Multiplier $\mathcal{L}$ that ensures the unbiased nature of the estimator and

$$\mathbf{b} = [\bar{\gamma}(d(\hat{\mathbf{x}}, \mathbf{x}_1)), \bar{\gamma}(d(\hat{\mathbf{x}}, \mathbf{x}_2)), \dots, \bar{\gamma}(d(\hat{\mathbf{x}}, \mathbf{x}_N)), 1]'$$

$\mathbf{b}$ vector contains set of theoretical semivariances from the nodes to the estimated point $\hat{\mathbf{x}}$.

Finally, solving the aforementioned System of Linear Equations SLE (54), the RSSI estimation is obtained. However, to determine the fingerprint update, the following decision rule has been implemented:

$$\frac{1}{N} \sum_{i=1}^N \|\hat{z}(v_i(\hat{\mathbf{x}})) - z(v_i(\hat{\mathbf{x}}))\| \leq \mu \quad (55)$$

where μ is an empirical threshold which value is in the range $\mu = (0, 3.5)$

The general outline for Kriging implementation is described in algorithm 3. In this algorithm, the sequential steps to obtain the RSSI estimation and to compare it with data gathered (RSSI from sensors) in order to make a decision on the Fingerprint update stage.

4.7 Experimental results

In this section, the experiments performed to assess MUFAF and the results obtained in comparison with existing methods are detailed. In the simulations, the laboratory scenario drawn in figure 20

Table 5: Average error comparison of MUFAF with several methods expressed in centimeters.

	XBee	CC2420	WiFi	Bluetooth
Euclidean Distance [86]	110.71	150.367	122.72	150.57
Mahalanobis Distance [86]	108.33	140.05	112.12	142.58
K-NN [87]	105.87	139.58	157.42	161.29
Gaussian Kernel [83]	67.27	61.92	63.30	170.34
MUFAF Exponential Kernel	63.05	59.49	79.73	137.15

is considered. This room represents a typical indoor environment with human activity, people flow conditions controlled and harsh propagation conditions due to multipath fading, obstacles due to furniture and wireless interferences due to electronic devices, further details can be found in [71]. Additionally, as described in section 4.2, four $\mathbf{v} = v_1, \dots, v_4$ nodes are deployed at the lab corners. Furthermore, every node is equipped with four sensors $l = 4$: Bluetooth, WiFi, and two 802.15.4 receivers (Xbee and TICC2420). Firstly, the TTP performance is assessed by comparing MUFAF to other methods that employ fingerprint technique. The experiment is run for a total of 1000 different routes like the one described in figure 20. Additionally, every route $\mathfrak{R}^l = \{\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_{\varpi=90}\}$ comprises 90 steps. A fingerprint database has been created with the cells distribution shown in the scenario figure 20. Measurements were carried out with the four sensors reading simultaneously in a period $\tau = 5$ minutes per cell.

Figure 22: Average error using both fusion architectures in path estimation for an experiment composed by 300 routes with 90 steps per route.

Table 6: Comparison of architecture error for 300 routes

	Track-to-track	Sensor Group
Average error cm	49.68	65.84

The average error for every method is detailed in table 5. MUFAF results are significantly better than deterministic methods. In addition, MUFAF results are better than Gaussian Kernel methods except for WiFi sensors. This is probably due to two aspects: (a) distances considered are not long enough to appreciate WiFi signal attenuation and (b) the histogram distributions are similar to Gaussian functions.

Furthermore, results obtained for the aforementioned sensor technologies are shown in figure 21. In this figure it can be observed that Bluetooth technology errors in estimation are larger than the other technologies. The reasons for this result can be the low power transmission of Bluetooth devices and interferences. Conversely, 802.15.4 based technologies perform with a mean error under 65cm

Moreover, the architecture patterns described in section 4.5 were also analyzed. The experiment consists of a set of 300 random routes with 90 measurements per route. Comparison results of both architectures are drawn in figure 22. It can be appreciated that TTTF outperforms KSGF in terms of error. TTTF involves to run TTP separately to fuse routes obtained for each sensor showing better tracking results than performing KSGF. As a consequence, TTTF allows optimization of KF parameters in every specific sensor. However, the average error for each architecture pattern is shown in table 6, where the difference is 16.16 cm which demonstrates that TTTF outperform KSGF.

Figure 23: Empirical threshold for decision rule in AFU. This parameter was obtained by performing iterative simulations for 300 routes with multiple μ values in the range (0,3.5).

Finally, the Adaptive Fingerprinting Update (AFU) was evaluated. In order to see the evolution of the distribution, an experiment consisting in an initial Fingerprinting containing only 30 measurements per cell is performed. Firstly, in figure 19 the evolution of histograms through AFU is illustrated. In top figures, the initial samples distribution is shown while in bottom figures the evolution of distribution

after 300 samples is displayed.

However, one of the most relevant aspects for AFU is the threshold definition. An experiment has been carried out to obtain the performance of AFU applying several values to μ . The results are depicted in figure 23, where the average error for 300 routes with fixed μ values show that the optimal value is $1.3 \leq \mu \leq 2$.

Figure 24: Error evolution for Position Estimation using AFU with a threshold $\mu = 1.5$.

Furthermore, taking into consideration the values obtained, a MUFAF testbed was performed with a predefined decision threshold $\mu = 1.5$. Fingerprinting database was initialized with only 30 samples and run our algorithm for 300 routes. The results are detailed in figure 24. In the one hand, blue line shows the fitting model that minimizes the minimum square error (MSE) of estimations for TTTF. On the other hand, red dotted line describes the model that minimizes MSE for KSGF. Results show that MUFAF employing TTTF in combination with AFU minimize the impact of changing conditions.

A particular example of routes and MUFAF performance is shown in figure 20. Green dotted line describes the real route where as the red line describes the path estimated for both: TTTF (left side) and KSGF (right side).

The entire dataset containing fingerprinting measurements for every technology, as well as example routes are totally available for testing purposes and comparison with the algorithms described here. This information can be download from <http://www.gatv.ssr.upm.es/~ghp/>

5 Chapter 5: Improving detection by using Deep Learning techniques.

Chapter Summary

This chapter provides an alternative perspective to the problem of person position estimation and tracking in two views: (a) On the one side, it exploits the high-performance of Deep Learning methods for the problem of estimation in fingerprinting. (b) It proposes a novel Deep-Learning method to improve the detection accuracy in each cell of the fingerprint. The idea behind is to attain a lower error at in-cell estimation. The contents of this chapter are related to the Q1 Ranked Journal “*SWiBluX: Multi-Sensor Deep Learning Fingerprint for Precise Real-Time Indoor Tracking*,” in IEEE Sensors Journal, vol. 19, no. 9, pp. 3473-3486, 1 May1, 2019 [39].

5.1 State-of-the-Art

In the recent years, scientific methods for data analysis of human activities have become a relevant topic to extract behavioral patterns [103, 86, 93] which represent valuable metrics for multiple purposes such as research, business, security, health [94] among others [86]. Indoor/outdoor tracking has attracted an extensive focus of research due to the continuous progress in computing capabilities and sensing technologies such as RGB, LIDAR, ultrasound, infra-red, etc.

An extended approach for people tracking is based on Computer Vision techniques using RGB sensors [94]. However, occlusions, lighting changes among others, constrain the estimation accuracy of these methods. Moreover, recent RGB-D and LIDAR sensors (i.e Intel Real Sense or Microsoft Kinect [95]) allow to obtain also depth information attaining a significant accuracy improvement compared to RGB estimation. Nonetheless, short detection range in addition to aforementioned issues decrease the application of these devices. There exist several works that employ ultrasound sensors [96] for tracking. However, lack of accuracy, difficulties associated to multi-object detection and coverage range as well, make these approaches unfeasible for scenarios where individualization is required (i.e places with limited capacity).

Mobile technologies have fostered the development of methods for position estimation and tracking [24] due to two main reasons: (a) these devices are widely used among population (smartphones) and (b) fulfill most of the aspects concerning privacy.

Advances have empowered personal devices with several wireless technologies (IEEE802.15.1 Bluetooth [119], IEEE802.11g/n WiFi [120] or IEEE802.15.4 [21]) that can be used for indoor position estimation. In particular, two metrics are used for the mentioned technologies: Time-of-Arrival (ToA) and Received Strength Signal Indicator (RSSI) [105]. The former relies on obtaining the time lag of a packet forwarded to an end-point and estimate the distance between devices. However, strong requirements in terms of hardware and synchronization make this proposal hard to implement. The latter approach is most extended as devices are able to obtain the RSSI parameter from packets although propagation effects such as multipath, fading, reflections among others rise a challenge for processing techniques applied [97, 99].

The inclusion of inertial sensors (accelerometers and magnetometers) has increased the set of features

Figure 25: *Left:* This figure represents the real monitored scenario and the cells distribution for fingerprinting. Five nodes equipped with the selected wireless technologies are deployed across the area (*circles*). Doors are drawn in red and furnitures in gray respectively. *Right:* the reference coordinates system for inertial and magnetic sensors as well as the allocation of devices employed are illustrated.

that can be applied to prediction models, yielding to meaningful accuracy improvements for person movement monitoring. For the aforementioned techniques (RSSI and ToA), a Wireless Sensor Networks (WSN) composed of a set of nodes deployed across the area under monitoring is utilized due to its low-consumption, self-organizing and easy deployment capabilities. Extensive research has been conducted in the literature to address the person tracking problem using RSSI [141, 86, 108, 104]. These proposals are mainly classified into deterministic or statistical methods. On the one hand, deterministic techniques adopt a theoretical model for propagation and geometrically combine the measurements gathered from several nodes to estimate the person position. On the other hand, statistical techniques, such as *Fingerprinting*, characterize the environment by creating a background dataset that contains the measurements at predefined locations. Several Fingerprinting parameters must be adjusted to achieve good performance in position estimation procedure. However, in most of the existing research, there is not a clear description neither of how these parameters were obtained nor how datasets are updated to face with changes in the scenario.

Machine and deep Learning algorithms provide novel tools to cope with large datasets, learning complex non-linear patterns from input information [106, 107]. Topics involve data science [109], image classification, detection and recognition [110] among others. Regarding the proposed scenario, Fingerprinting can be treated as a huge dataset of RSSI measurements where these algorithms can be applied [111, 112, 113]. Once the target position is estimated, tracking is performed by filtering these locations using well known methods such as Kalman Filters (and its nonlinear variants) and Particle Filters [127, 128, 129] using a reference movement model.

In this chapter, a complete framework for indoor estimation and tracking using wireless technologies and inertial sensors (WiFi, Bluetooth and Xbee technologies). It adopts two layers: (a) Fingerprinting for position estimation and (b) Neural Networks in combination with a novel Gaussian Outliers Filter (GOF) for person tracking. Highlighted contributions are:

- Describe a complete and comprehensive analysis of the most influential parameters in Fingerprint construction.

Figure 26: Example of RSSI variance boxplots in the collected Fingerprinting database is presented for each node in different cells selected for a particular wireless technology (XBee). Variance in measurements is large, yielding to significant fluctuations in the signal distribution over a particular location (i.e. a cell in the monitored room).

- Outline a procedure to establish boundaries in parameters selection.
- Draw a Deep Learning (DL) approach for Fingerprint employment in indoor estimation.
- Provide a complete overview of different methodologies in positioning and tracking.
- Present a disruptive and modular architecture for accurate indoor tracking based on a combination of statistical and DL methods.
- Propose a novel Gaussian Outlier Filter for Tracking refinement able to improve the Neural Network performance up to 10%.

The remainder of this section is organized as follows: in 5.2, the overall indoor tracking problem is formulated. Subsection 5.3 details the most influential aspects in fingerprinting technique for position estimation. The proposed system is outlined in subsection 5.4.

The statistical analysis can be further improved based on Fingerprinting methods that overcome the mentioned difficulties, allowing to collect enough data for indoor scenario characterization. Classical propagation models are commonly used, however, harsh and changing indoor conditions make these methods unfeasible due to its low accuracy. Other indoor positioning methods require to adjust several parameters subject to the monitored area, a radio map or additional measurements. Fingerprinting aims to construct a reliable radiomap with a few measurement points, enabling the use more complex techniques to better characterize the final estimated position.

Multiple works have been presented for position estimation using Fingerprinting technique, neglecting the details on how the fingerprinting dataset was calculated [23, 141, 98]. Nonetheless, some infor-

mation on its implementation was provided in [100, 101, 130]. In [101], several details regarding the propagation effects have been outlined. However, justification of how the cell number/dimension or device allocation for RSSI collection phase is not provided.

In [102], the effects of user body absorption, antenna orientation and Line Of Sight LOS/NLOS propagation are presented. Results showed how the RSSI attenuation is larger in the presence of human occlusions and it must be considered for Fingerprinting creation. In [103], a four-directions (oriented) Fingerprinting is drawn to enhance the dataset resolution as well as to mitigate the estimation error due to the person/device orientation. Furthermore, an analysis of cell dimension influence showed that large variance of RSSI measures affects the performance of positioning algorithms. These studies raised challenges on how to reach reliability of the RSSI measurements in Fingerprinting.

Probabilistic methods exploit the RSSI variance for position estimation [68, 69, 121]. Recent approaches based on Machine/Deep Learning techniques are capable of employing the RSSI dataset to learn the optimal model that relates the measurements with the final position estimated. In [124, 125, 126], methods such as K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Neural Networks (NNs) are applied. Moreover, works incorporating adaptive updates of Fingerprinting parameters to adjust the models to changes in environment conditions have been depicted in [157, 72].

Estimated locations are provided as output from the aforementioned techniques. However, these estimations are noisy and neglect temporal dependency. Therefore, using the positions from Fingerprinting stage, several filters can be applied on top to fit the routes obtained to a movement model as well as to reduce the impact from outliers. The most common approaches are Kalman Filters and Particle Filters [127, 128, 129] which are based on state-space models and combine measurements from different sources to estimate the next state of the system. In fact, Fingerprinting can become an important noise source if the parameters are not properly adjusted.

In the following works: [112, 113, 114], a complete survey of classical, Artificial Neural Networks (ANN) and Reinforcement Learning techniques, attaining a positioning error in the range 1-3 meters is presented. However, it extends these approaches by involving additional information extracted from inertial/magnetic sensors to improve the quality of the positioning and tracking algorithms. A custom feature vector is proposed as ANN input to tackle the problems related to Fingerprinting technique. An additional layer to enhance spurious outliers filtering is detailed. The results shown that its complete sequential architecture outperforms the literature.

5.2 Problem Statement

A typical indoor scenario is characterized in Figure 25 (scenario used in the experiments performed in this work) where the width and depth are denoted by W, D respectively. Additionally, a set of cells $\mathbf{c} = \{\mathbf{p}_l \mid l = 1, \dots, L\}$ with geometrical center at $\mathbf{p}_l = (x_l, y_l) \in \mathbb{R}^2$ is defined, where c_i is a particular cell. The distance from a cell to another is marked by $d(c_i, c_j)$ using euclidean distance as metric.

It is assumed that a set of nodes are deployed across the indoor scenario under monitoring. These nodes can be modeled as $\mathbf{v} = \{v_{i,\lambda} \mid i = 1, \dots, I; \lambda = 1, \dots, \Lambda\}$ and are fully connected. Furthermore, the nodes are located at Cartesian coordinates $\mathbf{p}_i = (x_i, y_i) \in \mathbb{R}^2$. Moreover, the nodes are equipped with sensors from wireless technologies λ able to extract the Received Strength Signal Indicator (RSSI) from a packet. In particular, sensors from three wireless technologies are considered: IEEE802.11.g/WiFi [120], IEEE802.15.1/Bluetooth v4.0 [119] and IEEE802.15.4/XBee. Let $RSSI_{i,\lambda}$ to describe the RSSI

Figure 27: An example of different orientations in the same cell is presented. The orientations have been concentrated in four angle ranges. The large variability in the measurements depends heavily on the device's orientation in each cell.

measurements gathered by node i with technology λ .

It is assumed that a/several individuals is/are moving in the scenario wearing mobile devices endowed with at least on of the aforementioned technologies (smartphone, wearable, etc). The main problem addressed in this work is the individual tracking using the WSN information.

Optimal cell distribution for Fingerprinting is one of the main issues addressed in this work. The most common aspects include: (a) identifying the most influential parameters for Fingerprint construction, (b) defining some approaches to mitigate the effects of cell distribution, (c) data fusion of several wireless technologies, (d) analyze and assess deterministic and statistic methods for position estimation based on RSSI dataset as well as Machine/Deep Learning approaches, (e) a simple outliers filter based on estimations given by the algorithms and parameters extracted from sensors and (f) the implementation of final tracking filtering algorithm.

Raspberry Pi devices were used as nodes, equipped with XBee S1 (protocol IEEE802.15.4, indoor range 30 meters, transmission power 0 dBm) Bluetooth USB Dongle PBU40 (protocol IEEE802.15.1 v4.0 BLE+EDR, indoor range around 10-20 meters, transmission power -3 dBm) and WiFi USB dongle TP-LINK TL-WN722N (protocol IEEE802.11.(b,g,n), indoor range around 40-80 meters, transmission power around 21 dBm), whereas the wearable was endowed with Inertial Movement Unit (IMU) Adafruit 9-DOF LSM9DS0 with accelerometer, gyroscope and magnetometer integrated.

The orientation of the wearable device attached to the body is presented in the right side of Figure 25. The device is allocated on the person hip and axes (X,Y,Z) in the figure match the IMU axes.

5.3 Fingerprint calculation

Aforementioned factors affect the quality of RSSI measurements yielding to large variance in the power level detected by every node, which is critical when the detection is performed in an indoor scenario as the one shown in Figure 25. Fingerprinting is employed to characterize the monitored scenario by splitting it into a set of cells and obtaining RSSI measurements at every cell for a predefined time. However, some important aspects such as the number of cells, the point of interest for cell measurements and the coverage area are generally neglected.

5.3.1 RSSI Variance

Initially, it can be argued that the larger the number of cell is, the higher characterization resolution achieved is. However, if the number of cells is large, the statistics could be similar between each other, increasing the estimation errors for Fingerprinting cells with identical behaviors or very close among them. Consequently, it can be also stated that the larger the number of fingerprinting cells is, the higher probability of error in estimation is.

Therefore, an analysis is required in order to select the cell size to prevent the overlap between RSSI measurements collected in the nearest cells. In practice, RSSI values overlap for different distances in multiple cells due to the propagation conditions. It can be inferred that the more distances with the same RSSI reading overlapping, the higher error probability in position estimation.

An example of this statement is depicted in Figure 26 where a set of four cells and all the RSSI measurements collected by all nodes are presented. A single node has a coverage range constrained for every particular technology, and any of these covers the whole area. As a consequence, unexpected effects such as a non-uniform data distribution from sensors around this position or even lack of coverage makes difficult to fit a proper propagation pattern for each cell. As an example, for Cell 8, it is shown that measurements collected from Node 2 and 3 (the closest nodes in our deployment) are larger than measurements collected from Node 1 and 5, whereas no measurements from Node 4 (the furthest node with more walls in between).

5.3.2 Wireless Technology Signal Attenuation

Wireless technologies used in fingerprinting have different signal distributions across the distance. There exist several approaches for indoor propagation modeling where the signal attenuation is relevant to define the cell size. As an example, cases where signal attenuation in a cell is very low (i.e. almost flat) have a similar distribution giving to unreliable estimations for this area. Based on this statement, the nodes must be allocated accordingly to the area under monitoring avoiding to have these nearly from each other. This effect is mitigated in the presence of multiple wireless technologies as its particular propagation patterns allow to distinguish the locations by fusing the data from each one.

5.3.3 Sampling Time per Cell and Synchronization

The data collection period for every cell is relevant due to the need of gathering enough RSSI measurements to obtain a representative sample. The number of packets forwarded depends on the processing

capabilities of transmitting devices. In order to synchronize the measurements collected, it is assumed that the transmitter device is able to embed the timestamp to the measures collected. Fixed length slicing windows of 40 ms were created to collect measurements that are assumed to be synchronized. In the tests performed in this work, the sampling frequency for the considered technologies is up to 25 samples per second and the sampling time per cell was adjusted to 5 minutes.

5.3.4 Object or Person Tracking

The absorption of people must be considered for human tracking applications. Several works have addressed this problem trying to model the human absorption [123]. Empirical tests were performed to determine that the average attenuation in detection due to human absorption, as well as measurements along several locations and different orientations with the device attached to the body and the device fixed without human interactions have been collected. In addition, some measurements with 3-6 persons moving along the monitored area were considered to emulate interference. The mean of the information gathered was used to establish an attenuation equal to 4 dBm.

5.3.5 Orientation

The antenna orientation effect has also implications on the statistical RSSI positioning algorithms. In general, such assumption remains only in the symmetric region of the antenna. However, when communications are presented in a non-symmetric region, which is generally the 3D case, there is a considerable variation of RSSI values. Moreover, the radiation pattern is not perfectly symmetric in all directions, having an impact on the signal propagation. An experiment rotating a 802.15.4 wireless device in the same plane, as explained in [122], has confirmed that the orientation of the device is a significant attenuation factor in the RSSI value. In Figure 27, it can be observed an example of the measurements collected in a cell rotating over all the angles and representing the RSSI measurements for a single node in four directions. There exists a large variability in the measurements when the device is pointing to the node (-30 dBm) or to the opposite side (-40 dBm).

5.4 System Architecture

Fingerprinting is a method for positioning in particular times, rather than a continuous tracking technique over time. Both, positioning and tracking methods are presented in the architecture drawn in Figure 28. The system architecture is composed by several stages: First, a synchronization process over the measurements is needed to create the feature vector. Second, a Deep Learning approach is used to perform the fingerprinting estimation. Third, a Gaussian Filter is proposed to reduce the outliers impact on the estimations. A weighted combination of estimations after the filter provides an estimated position which is inserted into a Particle Filter to track the person based on a realistic movement model. The use of additional sensors and the proper fusion of diverse sources improves the detection accuracy. The velocity estimation algorithm provides information about how fast a person is moving in the monitored area and the Madgwick [137] filter is used for orientation estimation to improve the final results in some stages of the algorithm.

Figure 28: System architecture is composed by several stages: *First*, a synchronization process and a Deep Learning approach is used to perform the fingerprinting estimation (green boxes) explained in Section VA. *Second*, a Gaussian Filter is proposed to reduce outliers in the estimations (orange boxes) presented in Section VB. *Third*, the step, velocity estimation and absolute orientation (blue boxes) are introduced in Section VC and VD respectively. *Fourth*, Particle Filter tracks the person based on a realistic movement model giving the final position (yellow boxes) introduced in Section VE.

Figure 29: General Scheme of the Training process. *From top to bottom:* Fingerprint creation from the N Network nodes for each particular technology in addition to the orientation values (Yaw) is depicted. The length of the feature vector inserted into the Neural Network is $(I(nodes) \times \lambda(technologies)) + 1(Yaw)$.

5.4.1 Fingerprinting Position Estimation

In this subsection, the literature methods described in section 7.1 are compared with Machine/Deep Learning approaches, where the output of this procedure is a probability for all the cells (at center) position based on the feature vector inserted to the algorithm.

$$[RSSI_{1,\lambda}, RSSI_{2,\lambda}, \dots, RSSI_{I,\lambda}] \quad (56)$$

Data in this format can be analyzed employing different methods such as Distance-based Metrics (i.e. Euclidean, Mahalanobis, Spearman), K-Nearest Neighbors including its versions (W-KNN) or more sophisticated algorithms such as Kernel Density Estimators. Nonetheless, these methods require some statistical information including both the first and second moments of the distribution as well as the percentiles of RSSI signals in each cell.

Machine/Deep Learning methods can exploit the entire dataset to learn more complex patterns using all the available information. The first step is to gather the RSSI measurements for each cell by every node and create a custom feature vector as shown in the Example 56. The resulting dataset containing these vectors for each cell is used to train the mentioned algorithms.

In Artificial Neural Networks (ANN) training stage, neurons are taught to identify specific incoming signals and whether activate or not for a particular pattern/output. In related works (see 7.1), the information of fingerprint in combination with the multiple modalities (from inertial sensors) were employed to train an ANN to learn the most likely cell locations of a target and generalize the probabilities of associated propagation effects. Once the ANN has been trained on samples of the known database, it is able to predict by detecting similar patterns for future input data even when these have not been seen previously. In this work, as input of ANN, the following feature vector (Equation 57)

Figure 30: Deep Neural Network composed of: an input layer with size the selected feature vector, two hidden layers and the last layer with size equal to the number of cells with a final Softmax activation.

to tackle the Fingerprinting problems in an indoor scenario is presented:

$$[RSSI_{1,X}, RSSI_{2,X}, \dots, RSSI_{I,X}, RSSI_{1,B}, RSSI_{2,B}, \dots, RSSI_{I,B}, RSSI_{1,W}, RSSI_{2,W}, \dots, RSSI_{I,W}, Yaw] \quad (57)$$

where $\{X, B, W\}$ indicates $X = XBee$, $B = Bluetooth$ and $W = WiFi$, I represents the number of nodes deployed and Yaw is the Heading angle estimated by the Madgwick Filter.

This vector includes all available wireless technologies which implies a synchronization stage that allows to employ data from each node deployed for each wireless signal. For instance, in a certain time slot, an XBee packet could have not being received due to the attenuation, whereas the rest of technologies collected properly the RSSI value. Moreover, in spite the Bluetooth signal is weak and the coverage range is not large, the estimation is possible since the rest of technologies have a larger coverage and might be available in cases where Bluetooth is not.

Furthermore, the Yaw/Heading angle introduced in the vector provides relevant information for the orientated RSSI in the cell as shown in Figure 27. The orientation affects measurements gathered, and these changes are registered during the Fingerprinting creation by a person wearing the device (to take in account body absorption impact) rotating continuously in the cell reference. The corresponding time to collect data on each cell is fixed to 5 minutes. In such period, the continuous rotation during the RSSI gathering allows to cope with multiple possible events to statistically characterize the RSSI cell distribution and the propagation behavior in each direction.

The selected feature vector is used to construct the complete Fingerprinting dataset for ANN training

Figure 31: Example of the Gaussian Outliers Filtering. The bidimensional Gaussian is centered at the previous estimated position and the deviation is the velocity calculated at the current stage. The cell estimated by the Neural Network, in this case, is wrong so the filter applies a low probability to it. In this manner, the larger probability is assigned to the nearest cells (the ones that the filter covers).

and testing. An ANN with two hidden layers is selected as network structure with a final *Softmax* activation (Figure 30). The activation function selected in the hidden layers is ReLu. By using two hidden layers, it is possible to represent an arbitrary decision boundary with rational activation function that can approximate any smooth mapping to any accuracy. Additional layers can learn more complex representations. Nonetheless, There is a trade-off between architecture layers (Deepness) and the computational complexity of the proposed system. In this work, significant improvements were not attained by increasing the number of layer (> 2) whereas the training process was longer and the delay in estimation was perceived.

Figure 29 presents the system complete training architecture. Synchronization between the wireless sensors employed is needed and the estimations of Yaw value from inertial and magnetic sensors during database collection step. The proposed scheme can be observed as a classification problem faced with an ANN where RSSI and inertial measures are inserted as input and the cell estimated is provided as output (classes).

Several approaches can be used to select the number of parameters to construct the network [140]. The method to choose both the number of hidden layers and neurons is the *Try and Error Method* that is characterized by repeated, varied attempts which are continued either until success or until the agent stops trying. A Forward propagation approach was selected to tune the network parameters as

Figure 32: *Left*: Three axis accelerometer measurements. Regarding the proposed referenced coordinate system it can be shown that the Y axis measurements are around $1g$ representing the gravity force ($9.8m/s^2$) in this axis. *Center*: Example of peaks/valleys detection. Red dots represent valleys and green dots represent peaks. It can be observed that the amplitude needs to be large between peaks/valleys to detect them correctly. *Right*: Estimated velocity after Monte Carlo Simulation to get the parameters for the movement model and filtering the output.

it works better in models with medium/low size of input data as the one presented in this work.

Moreover, weights initialization has played a crucial role in this work due to the fact that *zeros initialization* is heavily penalized during the training process by not learning at all during epochs with more complex optimization algorithms or learning quickly with simpler optimization algorithms. Therefore, *He initialization* [142] was used in this work.

Since the input data is not very sparse, some methods such as *SGD* (Stochastic Gradient Descent), *NAG* (Nesterov Accelerated Gradient) and *Momentum* are preferred. In the complicated task of selecting the optimal learning rate, *SDG* provided the best results on the dataset created in this work and it was chosen consequently.

The learning rate was fixed to a small value of 10^{-3} to find a local or global minima during the optimization by avoiding large jumps over the loss function surface. Further, no overfitting was observed in the experiments, therefore, no regularization method has been included during training.

For the training process, the data was splitted into two subsets, the training and test sets. The ANN model can learn more efficiently with low-range numbers as input. The proposed feature vector was normalized into zero-to-one values due to the fact that all these parameters are always between a range of -92 and -25 for RSSI values and yaw/heading value between $0 - 360$ degrees respectively.

Once the feature is passed, the back-propagation optimizes weights to minimize the *Categorical Cross Entropy Loss Function* using the selected optimization algorithm.

5.4.2 Gaussian Outliers Filtering and Weighted Position Estimation

Machine/Deep Learning based positioning algorithms provide each cell probabilities in the Fingerprinting distribution. In a real scenario, a person moving across the monitored area can stay at different

locations besides the cell center. The last layer in the Neural Network usually performs the *Softmax* activation function to obtain a vector with the cell probabilities for a particular input feature vector. However, in case of wrong estimations, a stochastic error is introduced in the next step of the algorithm using only the center of the cell as estimation.

Layer previous to the *Softmax* activation contains the probabilities for each cell of being selected for position estimation. The proposed system removes the *Softmax* activation layer (at testing stage) and applies a Probability Density Estimation (PDE) filter to separate the unlikely cells (i.e. it is not possible that a person jumps several cells from a previous instant). Additionally, this filter smooths the probabilities of the most likely cells to reach a fine tuning based on the last estimation.

The two-dimensional Gaussian Function (Equation 58) is applied over the outputs of the last layer in the ANN. The key concept is to multiply the outputs of this layer with the ones generated by the two dimensional Gaussian Function. General statistical tools, the *mean* and *deviation* on both spatial axis (x, y) are required to generate this function. The mean is obtained using the last position estimated (previous stage) by the Tracking Filter. The deviation is not a constant value and depends on the person movement. Therefore, an adaptive value is chosen to guarantee that the Gaussian Filter width fits with the person movement at each moment.

$$f_X(x_f, y_f) = \frac{1}{(2\pi)|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^\top \Sigma^{-1}(\mathbf{x} - \mu)\right) \quad (58)$$

Person velocity is the key parameter used as the Gaussian Filtering deviation due to its ability to adapt to person changes. To cope with multiple cases, the Gaussian Filter can be wide-spread allowing medium-far cells to be considered into the position estimation when the velocity is high or conversely, to employ narrow filters for slow velocity. This procedure changes the probabilities and provides a normalized vector **p(Gauss)**. Afterwards, a weighted combination of the cells based on these probabilities yield the final estimated location entering to the Tracking Filter as shown in Figure 31.

$$x_g = \sum_{l=1}^L \frac{p(Gauss)_i}{\sum p(Gauss)_i} x_l \quad y_g = \sum_{l=1}^L \frac{p(Gauss)_i}{\sum p(Gauss)_i} y_l \quad (59)$$

Equation 59 x_g, y_g shows the estimated position before the tracking filter and $p(Gauss)_i$ represents the probability of a cell (with center at location (x_l, y_l)) that can be candidate to the weighted combination.

5.4.3 Step detector and velocity estimation

In this work, the accelerometer measurements are used to extract information about the walking/s-standing state to estimate the velocity based on the signal analysis and patterns. The inertial sensors provide information about the acceleration in the three coordinates axes. An example of measurements gathered during a normal walking behavior are presented in left part of Figure 32. Root Mean Square (RMS) estimation from the three accelerometer axes for the signal analysis was used for processing the raw acceleration data and it is calculated as follows 60:

$$RMS_{acc} = \sqrt{a_x^2 + a_y^2 + a_z^2} \quad (60)$$

To detect steps, maximum (maxima) and minimum (minima) acceleration peaks can be determined by using top and bottom thresholds respectively. However, this method does not work properly due to the large variation of both maxima and minima amplitudes during the walking phases. This issue can cause either lack of peaks or multiple peaks in a single step period. Conversely, the maxima peaks have less amplitude variations than the minimum. To reduce or even eliminate completely this undesirable effect, a peaks/valleys detector with an adaptive threshold is employed. Such algorithm is known as Adaptive Step Jerk Threshold [118] and it is capable to modify the threshold to filter peaks or valleys with no sufficient amplitude or consecutive false peaks/valleys. Using this information, person step length can be determined.

There are empirical equations to obtain the velocity based on these parameters [131, 132, 133]. However, most of these approaches do not fit properly the velocity model analyzed in this section. The performance comparison allowed to establish that the best results are obtained using [131]. Equation 61 describes the velocity model based on the output of the peaks/valleys algorithm:

$$L_{step} = K_{conv} \sqrt[4]{A_{max} - A_{min}} \quad (61)$$

where A_{max} is a peak, A_{min} represents the consecutive valleys and K_{conv} is a constant for unit conversion. It is observed that the equation does not adjust this model perfectly due to the fourth square root and the unknown value for the conversion parameter, although this model presents a similar behavior. A Monte Carlo simulation was performed over the root value and the conversion parameter based on different walking measurements of diverse individuals minimizing the error between the real and the estimated velocity.

The last step is to filter the final estimated velocity values. An example of peaks/valey detection along with the estimated velocity by the algorithm over these values are presented in the center and right part of Figure 32.

5.4.4 Absolute orientation estimation

Common approaches in orientation measurement utilize low cost sensors that lack of accuracy. Nonetheless, the information of accelerometers, gyroscopes and magnetometers can be fused to attain reliable orientation estimations. Three often used choices include the complementary filter [134], the Kalman Filter [135] and the Mahony/Madgwick Filter [136, 137]. The Madgwick Filter was chosen due to its good performance when applied over non high precision sensors, low computational complexity and real-time processing.

There exist calibration algorithms to reduce the errors produced by sensors high sensitivity [138, 47]. In this work, the ellipsoid calibration algorithm [139] was selected because of its good trade-off between precision and data required for calibration. This algorithm collects measurements from all the possible sensor positions in the space to correct the bias as well as concentrate the corrected values in a normalized sphere of radius equal to one (Figure 33).

Once the calibration procedure has finished, the absolute orientation can be estimated. This filter uses a quaternion representation, allowing accelerometer and magnetometer data to be used in an analytically derived and optimized gradient-descent algorithm to compute the direction of the gyroscope

Figure 33: Magnetometer calibration results. *Left*: uncalibrated normalized data points, *right*: calibrated normalized data points after ellipsoid calibration.

measurement error as a quaternion derivative.

$$\begin{aligned}
 Yaw &= \arctan \left(\frac{2(q_1 q_2 + q_0 q_3)}{q_0^2 + q_1^2 - q_2^2 - q_3^2} \right) \\
 Pitch &= -\arcsin \left(2(q_1 q_3 - q_0 q_2) \right) \\
 Roll &= \arctan \left(\frac{2(q_0 q_1 + q_2 q_3)}{q_0^2 - q_1^2 - q_2^2 + q_3^2} \right)
 \end{aligned} \tag{62}$$

The filter output is the estimated quaternion (four dimensional vector) based on the actual sensors measurements. The quaternion vector is employed to calculate the Yaw/Heading, Pitch and Roll as shown in Equation 62.

Figure 34: Estimation accuracy in position estimation using several methods with different feature vectors in two main cases: The first case (*left*) corresponds to perform the normal Fingerprinting estimation based on the selected method, the second (*right*) shows the results after applying the GOF and the weighted combination proposed in this work. For horizontal axis X, B and W denote XBee, Bluetooth and WiFi technologies respectively.

5.4.5 Filtering Tracking with Particle Filter

The tracking algorithm is applied over the estimated position from the previous stage to filter and reduce estimation noise as well as to adjust the positions to an human movement model. It is applied to reduce the fast non-natural movement errors and to fit the model to a like-human movement as shown in Equation 63:

$$\begin{aligned} x_{pf} &= x_p + \hat{v} \cos(\theta) dt \\ y_{pf} &= y_p + \hat{v} \sin(\theta) dt \end{aligned} \quad (63)$$

where x_p, y_p are the coordinates of the previous positions estimated, $\hat{v}$ is the estimated velocity and θ is the yaw/heading angle. The employed tracking algorithm is the well-known Particle Filter [143] which is a common choice for applications such as robotics by estimating position, velocity and heading angle.

The idea behind Particle Filters is that any Probability Density Function (PDF from now on) can be represented as a set of samples (particles). Each particle has a set of values for the state variables representing arbitrary distributions, making it appropriate for non Gaussian, multi-modal PDFs. Therefore, it is feasible to find an approximate representation of a complex model (any arbitrary PDF) rather than an exact representation of a simplified model (Gaussian).

To start the algorithm, an initial belief state, denoted as $p(x_0)$, which is a rough estimation of the PDF

is required. For localization problems where no previous knowledge of the distribution is available, it is possible to scatter particles over the whole map. In this work, the particles are scattered around the first estimated position in the first stage with a Gaussian distribution centered at that position with deviation equal to twice the estimated velocity. This is done to cover a wide area of the map in case that the previous estimation might not be reliable. For each time step, the algorithm loop contains three main phases: *prediction*, *update*, and *resampling*.

In the *prediction* step, each particle is displaced according to the motion model. The resulting position from the motion model, will be somewhere in that cloud of particles. The resulting distribution of particles approximates the prior distribution.

During the *update* step, weights are assigned to the sensor measurements according to the probability of observing the sensor measurements from that particle's state to each particle. The weights ($\mathbf{w}$) are normalized and its sum is equal to 1 (Equations 64 and 65) to guarantee unbiased decisions.

$$\mathbf{L}_{xy} = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-\sqrt{(\mathbf{x}_p - x_g)^2 + (\mathbf{y}_p - y_g)^2}}{2\sigma^2}\right) \quad (64)$$

$$\mathbf{w} = \frac{\mathbf{w} \cdot \mathbf{L}_{xy}}{\sum_{i=1}^M w^{(i)} \cdot l_{xy}^{(i)}} \quad (65)$$

where L_{xy} represents the particle importance with respect to the previous estimation and M is the number of particles.

In the *resampling* step, a new set of particles is chosen maintaining the ones with the largest weights. Therefore, unlikely particles (small weights) at the fringe are not chosen whereas the most likely particles (near the center of the cloud) are replicated. As a result, the high-probability region has a high density and properly represents $p(x)$ (the posterior distribution). There are several methods for weights resampling [144], but the Systematic Resample [145] is selected as it faces with the event of particles wandering away from the point and its weight tends (or is equal) to zero. Consequently, a small number of particles implies a reduced number of particles contributing to the approximation of the distribution. Moreover, in order to determine whether resampling is needed or not, the coefficient of variation statistic [146] can be calculated using the following equation 66:

$$CV = \frac{1}{M} \sum_{i=1}^M (Mw^{(i)} - 1)^2 \quad (66)$$

The effective sample size can be calculated as shown in Equation 67. This parameter describes how many particles have an appreciable weight. Therefore, to check if resampling is necessary, the effective sample size can be tested against the number of particles as follows:

$$ESS = \frac{M}{1 + CV} \quad \text{Resample if } ESS < \tau \cdot M \quad (67)$$

where τ is a threshold between 0 and 1. In this work, this value is selected as 0.5 which means that not resampling is requested in case the half part of the particles exceeds the ESS value.

Table 7: Average error in position estimation (centimeters) for several methods using fingerprint at different stages of the proposed system architecture with the proposed feature vector

Positioning Algorithm Proposed System Stage	Fingerp	Fingerprinting After Gaussian Outliers Filtering	Fingerprinting + GOF + Weighted Combination	Particle Filter
Deterministic and Probabilistic Common Methods				
Euclidean Distance	259.56	NA	NA	156.34
Mahalanobis Distance	230.56	NA	NA	145.21
Weighted-KNN	210.56	NA	NA	135.67
Gaussian Kernel	184.56	147.78	120.54	102.53
Exponential Kernel	191.32	153.62	132.02	111.36
Machine Learning Methods				
KNN Classifier	135.99	111.21	101.77	87.08
Logistic Regression Classifier	133.76	111.32	99.89	82.32
Support Vector Machine	104.54	90.78	85.67	78.65
Decision Tree Classifier	97.87	90.62	84.98	75.76
Gaussian Naive Bayes Classifier	89.43	82.22	78.45	69.91
Random Forest	88.59	77.79	74.76	62.48
Gradient Boosting	85.32	74.65	71.78	60.56
Deep Learning Methods				
Deep Neural Network	80.67	68.72	61.88	45.41

Finally, the new estimated position is obtained based on the calculated weights and the particles after the prediction step as Equation 68 indicates:

$$x_f = \sum_{i=1}^M \mathbf{w} \cdot p_x \quad y_f = \sum_{i=1}^M \mathbf{w} \cdot p_y \quad (68)$$

where (p_x, p_y) are the x and y position of each particle respectively and M is the number of particles.

5.5 Results on Position Estimation and Tracking

This section shows results on location estimation with the most common methods in Fingerprinting positioning, comparing with the results achievable by different algorithms of both Machine Learning (ML) and Deep Learning (DL) techniques. Furthermore, results on the use of the Gaussian Outliers Filtering and the Weighted Combination of the outputs of this filter after ML /DL methods demonstrates an accuracy improvement in the position estimation. The final location of the person wearing the device is provided by using Particle Filter reducing the final estimated error.

The configuration of the scenario is shown in Figure 25 and represents a typical indoor environment where human activity can be observed and the people flow is controlled. Moreover, harsh propagation conditions arises due to the multi-path fading, the obstacles provoked by the furniture and the wireless interferences due to the existence of electronic devices. Further details can be found in [71]. Five nodes

Monte Carlo Simulation for Velocity Parameters estimation

Figure 35: Monte Carlo simulation over a range of possible values for adjust the the step length model to the detected peaks and valleys by the adaptive step jerk algorithm.

$\mathbf{v} = v_1, \dots, v_5$ are deployed and each node is equipped with three wireless sensors technologies $\Lambda = 3$: IEEE802.15.1 (Bluetooth), IEEE802.11.g (WiFi) and IEEE802.15.4 receiver (XBee).

The monitored area has been splitted into 70 cells of different size varying between 150×150 cm to 240×240 cm. The system is able to collect up to 25 RSSI measurements per second per technology. The collected dataset contains RSSI measurements from each employed wireless technology and Yaw values during 5 minutes in each cell. To consider scenario changes, the measurements have been collected in five different rounds of 1 minute per cell. After the synchronization stage, the database contains $7.5k$ feature vectors per cell.

To calculate the accuracy, it is necessary to know where the person is located during a complete route in the monitored area (a.k.a ground-truth). Several routes were collected to get the error in the estimations, comparing the estimated position at each algorithm step and getting the error from the estimated position over the real position using the euclidean distance metric.

Figure 34 shows the accuracy obtained using several methods for fingerprinting estimation taken different approaches for the feature vector in two cases. Accuracy is interpreted as the mean euclidean distance between the estimated position and ground-truth. In the first case, the fingerprinting stage is applied whereas the second case contains the accuracy after applying the proposed Gaussian Outliers Filtering in combination with the Weighted Combination of the probabilities for each cell. These feature vectors contain all the possible combinations of the three technologies employed.

Based on these two figures:

- It can be observed that all methods reach a larger accuracy in the estimations after the application of the GOF followed by the weighted cell combination based on the resulted probabilities.
- Different feature vectors present diverse results. As an example, Bluetooth technology alone

Figure 36: Particles (blue dots) managed by the Particle Filter at different moments during the tracking. The red dots represent the estimated position at each moment based on the particles whereas blue line depicts the final path estimated using the proposed architecture.

has the largest error in all methods as well as it is clear that the combination of technologies outperforms single technology-based estimation. WiFi technology provides a more robust RSSI (with less variance). Thus, all methods work well using this wireless technology in the feature vector.

- The use of the proposed feature vector outperforms the rest of combinations due to the introduction of the Yaw parameter which provides information about the orientation, improving the pattern learning process by Machine/Deep Learning algorithms.
- Deterministic and Probabilistic methods have the worst results compared to learning algorithms. The method including the feature vector proposed in this work can not be applied to distance methods such as Euclidean, Mahalanobis and W-KNN as enough relevant information can not be obtained from the Yaw to be applied.
- The Deep Neural Network achieves the largest accuracy results in both schema. Deep Learning algorithms can learn more complex and non-linear patterns that are limited in the Machine Learning methods.

The velocity model selects the parameters that reach the best fitting to the real measurements with the proposed system. A Monte Carlo simulation over a range of possible values was used to search for the best parameters in several velocity measurements. Finally, the mean of the parameters obtained at each evaluation over real measurements and the final equation are (Figure 35, Equation 69):

$$K_{conv} = 0.81 \quad r_{exp} \approx \frac{1}{3} \quad L_{step} = 0.81(A_{max} - A_{min})^{\frac{1}{3}} \quad (69)$$

Table 8: Different Neural Networks configuration presented with several assessment parameters

Eval. Parameters	Precision	Recall	F1 score	Time (ms)
One Hidden Layer 256 neurons				
Neural Network 1	78%	80%	79%	0.896
Two Hidden Layers 256, 64 neurons				
Neural Network 2	89%	88%	87%	1.235
Two Hidden Layers 512, 256 neurons				
Neural Network 3	90%	81%	85%	2.146
Three Hidden Layers 512, 256, 64 neurons				
Neural Network 4	93%	72%	83%	3.689

These parameters adjust adequately to the proposed method for velocity estimation in different cases when a person is walking slowly, rapidly or even when a person is stopped.

Analyzing the results presented in Table 7, it can be observed a significant improvement after each stage in the system architecture:

- GOF improves the final accuracy by filtering critical outliers that can degrade the performance of the final estimations.
- The Weighted Combination based on the resulted probabilities enhances the estimated position selection, not always placed at the center of the cell, thus reducing the final error.
- Particle Filter tracking algorithm in combination with the estimated parameters (Yaw, velocity) achieves a satisfactory performance with better estimated positions and reducing the average error up to 15 cm.
- Figure 36 represents an example of a person tracked and the particles used by the Particle Filter at three different moments. The number of particles is decreasing over the time according to the selected threshold by the *ESS* parameter. It can be shown that the inclusion of both Velocity and Yaw provide meaningful information to the filter achieving high accuracy in the estimations.

The ANN has been tested with different configurations looking for high precision while avoiding overfitting. Furthermore, the selection was subject to estimate quickly the position in order to maintain estimations in (near) real time. Four experiments were performed and are presented in Table 8. The number of epochs in all configurations is 100. The batch size has been reduced to 64 to learn quickly reducing the overfitting at early stages of the algorithm. Early stopping is used to finish the training if the precision tends to decrease due to overfitting. To ensure a correct dataset splitting in training and validation sets, at learning phase each network was trained ten times with random splitting at each training and the final precision/recall/F1 values are the mean of these ten experiments. The training data size after each splitting is the 85% of the total dataset collected and the test/validation data has the 15% of the total measurements.

- Configuration four (in table 8) reaches the best precision but the recall is not large. This is a symptom of the starting of overfitting since the number of neurons in the network layers is large

Figure 37: Training Loss and Accuracy for Training and Validation sets in two hidden layers (256, 64 neurons respectively) neural network configuration.

regarding the low size of the input feature vector. It means that the ANN learns quickly the patterns and after some epochs, it *overfits* the final estimations reducing the performance.

- The first configuration presented a short processing time and an acceptable precision, although it is lower than precision obtained by configurations two and three.
- Configurations two and three attain the best precision and recall results. However, configuration two has been selected for the experiments due to its reduced computation time.

Configuration two has been selected for the experiments regarding the precision of the estimations and the computation time (a simple schema of this network is presented in Figure 30). Both precision and loss during training are depicted in Figure 37. The loss value decreases after each iteration in the training set and the loss in the validation set follows the same decreasing pattern. The precision grows up in training as depicted in the figure but early stopping ended the training process at epoch 49 as the next iteration precision and loss do not decrease.

The entire dataset containing fingerprinting measurements for every technology, as well as example routes are totally available for testing purposes and comparison with the work described here. The scripts for parsing the data are provided and the guidelines in the README file. This information can be downloaded from <http://www.gatv.ssr.upm.es/~abh/>

6 Chapter 6: Improving the quality of measurements estimation by using Data Augmentation.

Chapter Summary

This chapter cover two main topics: (a) The use of Generative Adversarial Networks (GAN) models to increase the amount of Fingerprinting data, by performing data augmentation and (b) The use of GANs to speed up the training in image generation. These topics were included in the same chapter as rely on similar type of networks. The former topic is related to contents of the papers: “*Recurrent model for wireless indoor tracking and positioning recovering using generative networks*” [40] and “*Training speed up in Generative Adversarial Networks by Gossiping Feature Maps*”.

6.1 A Novel Approach for Indoor Tracking Using Adaptive Fingerprint Estimation Generative Models

6.1.1 State-of-the-Art

The relevance of this topic relies on the fact of allowing to analyze and characterize the behavior of objects and humans in an indoor environment [160]. As an example, this topic will allow to profile clients behavior in places to optimize the processes and services, or even better resources/products allocation and distribution. As aforementioned, human recognition and monitoring services are relevant for proper integrated health-care implementations where activities can be classified into multiple classes according to people evolution patterns [3, 4, 5]. In industrial applications, the indoor tracking information is meaningful for operations traceability, quality assessment, among others, yielding to direct cost savings. Airports flow optimization and monitoring is a nowadays an attention focus where metrics of multiple detection, and tracking might contribute to provide the services in an efficient manner [162, 163]. A very extended application is given by costumers habits shaping for smart selling in retail.

In general, indoor positioning and tracking methods heavily depend on the application requirements in terms of precision, capacity, scalability, systems’ time of response, and more recently, privacy and data protection [164, 165, 166]. Nowadays, there is a continuous and fast evolving and penetration of Internet of things (IoT), the advanced capabilities and low size of sensing technologies and the improvements of processing methods. These aspects endow the topic with a considerable dynamism and rise a broad number of opportunities to cope with the demanding application needs.

Alternatives are mainly based on diverse radio spectrum bands sensing, including acoustic, depth, infrared among others. In these schemes, a set of anchors are deployed across the area under monitoring. The aiming is to detect signals transmitted from devices attached to the target, and cooperatively to estimate the location. Multiple quotidian technologies such Bluetooth, WiFi, IEEE802.15.4 are employed to take the most of high penetration of such devices (smartphones). Recently, Radio Frequency Identification (RFID) is used in this topic due to the ability of detect events with passive tags.

Additionally, from a privacy, data protection and related aspects, the solutions can be classified according to the degree of intrusiveness. Camera systems generally have multiple privacy considerations

which make these approaches non-applicable in multiple scenarios.

Non-intrusive technologies rely on the wireless detection through parameters such as Time of Arrival ToA or Received Signal Strength Indicator RSSI. The former metric has strong requirements in terms of hardware and synchronization whereas the latter is more extended due to its simplicity and the large amount of devices supporting it. Furthermore, there are multiple issues associated to these technologies that must be considered. First, short range coverage can constraint its application. Second, due to propagation effects involving shadowing, multipath, fading, among others, the signal propagation is harsh to model and demands a huge effort from research community. Third, the environment conditions are complex, therefore, the models must be able to capture time variations. Finally, aforementioned issues directly impact on the systems' detection accuracy.

Moreover, the fast rise of Machine Learning and Deep Learning application to the indoor positioning and tracking topic, Artificial Intelligence based proposals are becoming popular. Different to classical statistical position estimation methods, Deep Learning (DL) approaches can better extract representative patterns for estimation and even adapting to changes in scenarios condition.

Traditional positioning methods are ingested with input signals to: (a) extract a set of features, (b) select the most significant features and (c) perform a prediction/classification task. The feature extraction process aims at transforming the input signals into a domain where significant patterns can be detected. According to the variation observed, the diverse features are weighted to assess its influence in the decision task and finally, traditional classification / prediction techniques are used for decision making.

As aforementioned in previous chapters of this thesis, the Fingerprinting methods have become more popular with respect to the traditional classical propagation models.

Therefore, the efforts are being devoted to find the most efficient manners to construct the Fingerprint database. In previous chapters, the entire procedure to build the fingerprinting was detailed, as well as the main issues and challenges. Nonetheless, several approaches involving compressive sensing to retrieve fingerprints where estimations are not available [162, 167]. The idea behind of these papers is to use some nearby locations as reference to estimate the value at unknown positions. Moreover, in [168], an initial scanner input is used as reference to adjust a large portion of the fingerprint measurements, reaching a low overhead. Moreover, Gradient applied over the RSSI is used as metric for the fingerprint database creation in [169]. Then the gradient is applied to estimate the positions nearby of the reference's one.

A novel technique was proposed in [170], where WI-FI APs positions were detected using the IEEE Zig-Bee signals. In [171], a similar approach to the one presented in chapter 4 of this thesis to construct the fingerprint was presented. This proposal is used to then estimate the value of RSSI using distributions similar to Chapter 5. However, these approaches neglect the temporal factor. Additionally, the use of NN as described in previous chapter outperforms the results reached in these papers.

Furthermore, [172] proposed a system for automatic construction of fingerprint maps for device-free WLAN using generated deterministic and probabilistic approaches. In [173] Matrix Completion framework was used to create the complete training maps, by using only some selected samples.

In [174], A Radio Frequency Identification (RFID) system attached to a person was used to update the Fingerprinting databases. moreover, in [175] an Enriched Training Database (ETD) as a service was proposed. It provides multiple functionalities to perform data augmentation using Generative Models,

and allowing simulations of routes creations based on predefined trajectories. Similar works, that use deterministic approaches for the path estimation and generative models for the data fingerprinting exist in the literature [176, 177].

Probably, the paper most similar to the work presented here is [178], where data is sampled at all the reference points, and then, more data is generated with an amplitude feature AF-DCGAN models, with which can effectively expand the number of samples in fingerprint database.

Deep Learning approaches can be classified into supervised, unsupervised and semi-supervised. In this work, supervised techniques will be adopted for adaptive fingerprint estimation. It is composed of three main steps: *(a)* signal preprocessing includes an adaption of inputs to the system involving flattening, normalization among others. In *(b)* Training phase, the network iterates to learn over the dataset to learn patterns (update weights) associated to an output subject to a loss function, and *(c)*, creates a model that can be loaded and employed to predict and classify incoming data. Therefore, the main challenge is to develop non-intrusive based solutions for multiple target detection and tracking with a high accuracy and a short processing time.

Main Contributions

The main contributions of this section are stated as follows:

- A complete study of the most influential parameters in the indoor environment tracking
- A dataset containing multiple cases stated here and a reference case of study to allow scientific community to compare their results with the ones proposed in this work.
- A complete framework based on Fingerprinting technique in combination with Deep Learning DL Architectures to face a continuous positioning and tracking problem.
- A Generative Adversarial Network model to adaptively create samples according to the environment conditions.
- A repository to access datasets, parsing utilities, models and code.

In subsection 6.1.3, the overall technical approach is presented. Subsection 6.1.4 details the experiments performed to validate the solution proposed. The discussion on results and causes is drawn in subsection 6.1.4

6.1.2 Problem Statement

In this subsection, the main assumptions of this work are described. Algorithms and methods validation were performed in the same scenario than previous chapters and depicted in figure 15. This indoor scenario emulates the real conditions of an office with people flow and condition changes (lighting, propagation, etc...). The room size is $5m \times 9m$ and $2.5m$ height. The hardware infrastructure is composed by a set of 4 Access Points APs allocated at top room corners.

It is assumed that a set of nodes are deployed across the indoor scenario under monitoring as depicted in figure 38. These nodes can be modeled as $\mathbf{v} = \{v_{i,\lambda} \mid i = 1, \dots, I; \lambda = 1, \dots, \Lambda\}$ and are fully

Figure 38: Images of the fingerprint measurement setup

connected. Furthermore, the nodes are located at Cartesian coordinates $\mathbf{p}_i = (x_i, y_i) \in \mathbb{R}^2$. Moreover, the nodes are equipped with sensors from wireless technologies λ able to extract the Received Strength Signal Indicator (RSSI) from a packet. In particular, sensors from three wireless technologies are considered: IEEE802.11.g/WiFi [120], IEEE802.15.1/Bluetooth v4.0 [119] and IEEE802.15.4/XBee. Let $RSSI_{i,\lambda}$ to describe the RSSI measurements gathered by node i with technology λ .

In the experiments performed, both, open-source software and hardware technologies were employed in this work. Two main components were defined for Access Points (APs) and sensing devices respectively. The former type was composed by 4 Raspberry PI V3¹⁰, deployed at top corners of the area under monitoring, whereas the nodes were based on arduino technology¹¹. The wireless technology chosen is IEEE802.15.4 wireless receivers (Xbee)[21]. Each device was equipped with a transmitter/receiver.

The problem addressed in this section is to develop a system that can create a set of measurements $RSSI_{i,\lambda}$ for each particular cell i using the existing fingerprinting database and the information learn by Generative Adversarial Networks.

6.1.3 System Architecture

Motivation

The Generative Adversarial Networks have become in perhaps, the most relevant research topic in the recent years, due to the wide range of applications in image, voice and video modalities. These Networks consists in the use of two Neural Network backbones trained for opposite tasks. The former backbone, called Generator $G(\cdot)$, takes as input a random noise vector and it is trained to generate an output (Image or video-frame, sound etc). Let's denote the output as $X_{fake} = G(z)$. Moreover, the Discriminator backbone, denoted as $D(\cdot)$ is trained to decide whether an input is fake or true. For this purpose, it receives as input two types of items: a real object (image ...) to be trained in what ground-truth is. The second type of data is the Generator's output $G(z)$. Then in the training phase, $D(\cdot)$ will learn the features of real items to make binary decisions whenever a Generator (*fake*) image comes in whereas $G(\cdot)$ will try to cheat $D(\cdot)$. Convergence in GAN training is reached when

¹⁰Resource available at <https://www.raspberrypi.org/products/raspberry-pi-3-model-b/>

¹¹Resource available at <https://www.arduino.cc/>

Figure 39: General architecture of a AC GAN

$D(G(z)) = 0.5$ which means that it is not able to distinguish whether an image generated by $G(\cdot)$ is real or not.

Of course, the process requires certain mathematical conditions to be stable. In concrete, the GAN stability relies on the Nash Equilibrium, having the following Loss function error:

$$L = \mathbb{E}[\log P(S = real \mid X_{Real})] + \mathbb{E}[\log P(S = fake \mid X_{fake})] \quad (70)$$

which can be expressed in a different way:

$$\min_G \max_D V(D, G) = \mathbb{E}[\log D(x)] + \mathbb{E}[\log(1 - D(G(z)))] \quad (71)$$

First, Discriminator maximizes the log-likelihood probability, which is the first term in equation 71. The Generator tries to minimize the second term in the equation.

Additionally, the information for both backbones can be inserted in supervised learning tasks by appending to z in Generator side, the label of the element inserted to not only learn features, but also classes associated.

As a consequence, both modules improve its performance: on the one side, the training of the Generative side is faster and obtains more information to reach better outputs. On the other side, the Discriminator backbone forces the Generator to learn the proper features of the inserted data.

Therefore, the notation is modified as follows: The input to the generator is not only the random variable, but also the label, denoted as c for the class. Giving as “fake” output $X_{fake} = G(c, z)$. Furthermore, the output of the classifier is also modified to classify not only whether the image/object is true or not but also the class that image/object belongs to.

Consequently, the loss functions can be updated as follows:

$$L_S = \mathbb{E}[\log P(S = real \mid X_{Real})] + \mathbb{E}[\log P(S = fake \mid X_{fake})] \quad (72)$$

which can be expressed in a different way:

$$L_c = \mathbb{E}[\log P(C = c \mid X_{Real})] + \mathbb{E}[\log P(C = c \mid X_{fake})] \quad (73)$$

Where equation (72) remains as the traditional *Loss Function* of GANs, where *Binary Cross Entropy* BCE among others are commonly used to measure quality of the real images compared with the fake or generated image.

Furthermore, the classes are also used in the loss function as shown in equation (73). Traditionally, Cross-entropy functions are employed in *loss function* training.

There is a general assumption when dealing with Generative Models. Ideally, *pooling methods* are not used [182]. The main reason is that in any of its versions (min, max or average), pooling information can not be recovered in the mirroring side (Generative backbone), making impossible to retrieve exactly the same image.

Figure 40: RSSI Evolution for a static point observed from multiple Access Points for a period of 24 hours.

Layer (type)	Output Shape	Param #
dense_55 (Dense)	(None, 256)	28928
leaky_re_lu_37 (LeakyReLU)	(None, 256)	0
dense_56 (Dense)	(None, 64)	16448
leaky_re_lu_38 (LeakyReLU)	(None, 64)	0
dropout_10 (Dropout)	(None, 64)	0
dense_57 (Dense)	(None, 1)	65
Total params: 45,441		
Trainable params: 45,441		
Non-trainable params: 0		

Figure 41: AC-GAN architecture for the Discriminator $D(\cdot)$ in the fingerprinting samples creation

Architecture

The architecture relies on the Auxiliary Classifier GAN, as depicted in figure [179]. Both GAN backbones $D(\cdot)$ and $G(\cdot)$ are described in figures 41 and 42 respectively. The former presents the Discriminator Backbone, which is formed by an initial dense layer with a 256-elements output ($dense_{55}$). It is followed by a Leaky activation function [180]. The second layer receives the “activated” output from the leaky function to provide a 64-element’s output ($dense_{56}$). Subsequently, an additional leaky layer is appended. Finally, a *dropout* layer [181] was added before the softmax layer to provide the binary decision on real/fake image (RSSI map).

Moreover, on the Generative backbone 42, the Network starts with the input, that in the present work was a *Random Variable* vector of size $N = 72$ with a 1-hot array representation of size equivalent to the number of fingerprint cells (in this work 28) ($dense_{58}$). This vector was inserted into a dense layer which provides a 64-size element output. Afterwards, the leaky function was applied in combination with a Batch Normalization layers. This was followed by a new dense layer which provides a 256-elements output ($dense_{59}$). Similarly, the leaky and batch normalization layers were appended to it. Finally, a dense layer followed by a reshape function provide the output desired (In this case a value of RSSI for every cell in the fingerprint (28) for each Access point (4)).

6.1.4 Experiments

This section provides the details on experiments performed. It is divided in two subsections, the former provides the details on how fingerprint dataset is created, as well how it varies in time. The latter, describes the data augmentation based on the data available and how the quality of positioning

Table 9: Experiments performed for each Dataset using the literature methods and the proposed GAF-GAN

	First Simulation	Second Simulation
Epochs	10000	30000
Batch Size	1290	1290
Sample Interval	100	100

estimation is increased by applying this technique. First, in order to demonstrate the variability of RSSI signals, an initial experiment, consisting in maintaining a set of nodes transmitting for a continuous period of 24 hours is depicted in figure 40. Although there exist some periods of continuous (at least in an statistic tool manner such as mean)

Fingerprint Creation

Fingerprinting is the most extended method for surface characterization in indoor model topic. It has been widely studied and consists in splitting the area nuder monitoring into a subset of cells where receivers are placed to collect data for a predefined time period. The idea behind is to model a medium-large area as the concatenated sum of small cells.

As mentioned in previous chapters, the steps followed are the following:

- To split the area into a number of spatial cells (28 in our experiments)
- deploy a set of sensors at each cell center location to collect information that will characterize the signal propagation across this area from all APs
- Store the information for post-analysis.

Fingerprint Reconstruction using AC-GAN.

Once a basic fingerprinting AC-GAN architecture is described in the previous subsection. Then, the “*Data Augmentation*” procedure can be performed. The architecture employed was described in previous subsection.

Two simulations were performed. The first one, run for 10000 epochs, with a convergence problem. Therefore, a second experiment was run with a total number of 30000 epochs, where the convergence is appreciated and the results are meaningful in terms of realist appearance compared to the real dataset.

In the first iteration, the GAN was not able to reach convergence due to an over-fitting. This phenomena arises whenever the network parameters are not settled properly or the dataset employed is not big enough, then the network learn the parameters for this dataset, but not the general behavior of the data (only the dataset, but not patterns are learned). The NN framework employed in these simulations was KERAS¹². Furthermore, the optimizer used was Stochastic Gradient Descent

¹²KERAS framework: <https://keras.io/>

Layer (type)	Output Shape	Param #
dense_58 (Dense)	(None, 64)	1344
leaky_re_lu_39 (LeakyReLU)	(None, 64)	0
batch_normalization_19 (Batch Normalization)	(None, 64)	256
dense_59 (Dense)	(None, 256)	16640
leaky_re_lu_40 (LeakyReLU)	(None, 256)	0
batch_normalization_20 (Batch Normalization)	(None, 256)	1024
dense_60 (Dense)	(None, 112)	28784
reshape_10 (Reshape)	(None, 28, 4)	0
Total params: 48,048		
Trainable params: 47,408		
Non-trainable params: 640		

Figure 42: AC-GAN Architecture: Generative backbone $G(\cdot)$ back for fingerprinting samples creation

The learning evolution can be directly observed in figure 6.1.4 for multiple epochs. In the vertical and horizontal axis, a division on the fingerprint cells is observed whereas the color represents the RSSI value. On top left side, it can be observed the original training image for a particular cell. In the subsequent figures, it is shown the training results for multiple epochs. As it can be observed, the evolution into a similar shape is improving directly related with the number of epochs in training.

The results showed that, in the initial epochs, the quality of the images created is not good (similar to the expected results (real image in top left)). However, the larger the number of epochs is, the better the quality of images generated. As we are dealing with generative models that try to characterize signals that have not a static behavior, there is a high difficulty in finding a metric that could cover appropriately the creation according to the typical values of RSSI. It is important to highlight that these values change depending on the scenario conditions. Therefore, the graphical representation with the red color representing the larger values (RSSI around 0) whereas the blue values present far away than the Access Point (lower than -65). To show the quality of the improvement, a small experiment for TPE as described in section 4 of this thesis was performed. Results are shown in 43. The star located at cell 10 was the fixed position where are red dots marked the estimated position, being cell 6 the one with the largest number of estimations.

Finally, the performance metrics of the network proposed are presented in figure 6.1.4. As it is observed, the convergence is not fast, however, the range of estimation allows to distinguish that with a reasonable number of epochs (≥ 30000) it is possible to obtain an improvement on the position

Figure 43: Example of the Position Estimation for a target located at cell number 10. The estimations were performed in cells 6 and 7 respectively.

Figure 44: Main evaluation metrics for the experiments proposed with the Generative Adversarial Networks: *(left)* Accuracy and *(right)* Loss

Figure 45: Main evaluation metrics for the experiments proposed with the Generative Adversarial Networks: *(left)* Accuracy and *(right)* Loss

estimations by the data augmentation previously performed.

6.2 Training speed up in Generative Adversarial Networks by Gossiping Feature Maps

In the previous section, an approach to generate new samples from existing dataset was presented. As it was shown, the GAN training is not a trivial task and in general the Generative Networks have multiple problems associated that have a direct impact on its performance. As a result, undesirable effects such as over-fitting or non-convergence can be presented.

In this subsection, some of these effects are outlined and method to speed up the training process is proposed.

Motivation

As mentioned in previous subsection, the GAN stability relies on the Nash equilibrium condition 73. However, this can not be always guaranteed as the design of the network, training parameters, dataset biased among other variables make difficult to reach the optimal values.

Some of the most common problems in Generative Adversarial Networks (GAN) training are:

- *Mode Collapse*: Generally, due to problems in the dataset (biased or limited samples size).
- *Non-Convergence*: The model oscillates, which is a sign of non-proper parameters setting.
- *Unbalanced backbones*: The Generator and Discriminator backbones must be balanced to avoid over-fitting.
- *Diminished gradient*: The discriminators rejects all generated samples, and the Generative side is not able to learn parameters (also known as vanishing).
- *High Sensitivity parameters selection*: Learning rate, optimization selected rates can be or very short step (causing slow learning) or large steps (causing over-fitting).

6.2.1 Proposed System

In this work, a method to speed up the GANs training process is presented. This method is depicted in figure 46 and it can be stated as follows: Let's assume a GAN network where the total number of layers ℓ is denoted by n where $\ell \equiv \{\ell_1, \dots, \ell_i, \dots, \ell_n\}$. The idea behind is that for each layer ℓ_i , the feature map for classification D^{ℓ_i} can provide ("gossiping") some information on the image characteristics that can be used by the Generator $G^{\ell_{n-i}}$, in layer ℓ_{n-i} for a fast learning. As observed in figure 46, The information of layer D^{ℓ_2} can be used by Generator in layer G^{ℓ_3} to learn the data from the real images.

The key factor is to determine how information from real discriminator images is combined with the data output obtained at Generator layers. In this work, it is proposed a weighted sum of each Feature Map. For this purpose, we define two weighting unities: (a) The creativity index, denoted as a , and

Figure 46: Gossip Adversarial Features GAN Architecture. The dotted lines indicate that the Feature Maps from (D) in this layer are passed in a weighted manner to the Generative side (G)

Speed Up Training Algorithm

Feature Extraction

- 1: Start with Discriminator $D(\cdot)$ over Real Images $\mathcal{R}$
- 2: in every single layer ℓ_i , to use the feature map and store it in a temporal variable $\hat{D}(\ell_i)$

Generation Speed Up

- 3: Afterwards, proceed in the Generative $G(\cdot)$ side.
 - 4: Combine the Feature map with a weighted version of the
 - 5: Finally, delete $\hat{D}(\ell_i)$
-

the gossiping index, represented by b . To guarantee unbiased and fulfill with the GAN equilibrium, the sum of both indexes must be equal to 1:

$$a_i + b_{n-i} = 1, \forall i \in \ell \quad (74)$$

From a global perspective, it might be argued that the influence of gossip Feature Maps shall be higher at the beginning of training and be reduced as far as G is progressively learning features of the images to be created. Consequently, the trend of a must be always decreasing.

Conversely, b , which controls the $G(\cdot)$ creation of images, can be not significantly large, but gradually it shall tend to 1. In fact, by the end of training, it must be fulfilled that $b = 1$.

Therefore, we can define the conditions of our parameters:

$$\begin{aligned} a + b &= 1, \forall i \in \ell \\ a &\rightarrow \downarrow \\ b &\rightarrow \uparrow \\ b_{Epochs} &= 1 \end{aligned} \quad (75)$$

$$a_{Epochs} = 0 \quad (76)$$

where *Epochs* represent the final epoch (or end of training). With this three conditions, multiple functions can be tested in the GAF-GAN scheme. In concrete, two type of functions were employed in this work: (a) *Exponential* and (b) *Linear* functions. In the next section, the details on the architectures, data-sets, parameters and results are provided.

6.2.2 Experimental Results

In this subsection, the multiple experiments performed are described. This work adopted as basis, the remarkable implementation of the diverse GAN architectures [183]. In this work, the most relevant GAN architectures are reviewed, as well as a outstanding implementation in both *Tensorflow*¹³ and *Pytorch*¹⁴ frameworks.

In this work, the Pytorch collection of GANs is employed. This repository contains the code for diverse implementations. This work was extended by:

- Creating implementations for each GAN, with is particular “*Gossip Adversarial FeatureMap*” version. (i.e. GAF-GAN, GAF-ACGAN, ...).
- Extending the scope of these implementations by appending a set of scripts to also apply the mentioned networks to the COCO Dataset¹⁵.
- Provide pre-trained models for all of the architectures presented.

¹³<https://github.com/hwalsuklee/tensorflow-generative-model-collections>

¹⁴<https://github.com/znxlwm/pytorch-generative-model-collections>

¹⁵COCO Dataset, available at: <http://cocodataset.org/>

Datasets

The following datasets were employed in this work:

- MNIST.¹⁶
- Fashion-MNIST.¹⁷
- CIFAR10 Dataset.¹⁸
- CIFAR100 Dataset.¹⁹

Diverse architectures were tested. however, the model proposed for MNIST, Fashion-MNIST, and CIFAR10. Network Architecture is exactly same as in infoGAN [184]

In general, all implementations with the same backbone have the following structure for the Generative side:

1. a Fully Connected Layer (FC) with output 1024 elements.
2. a Batch Normalization layer
3. a ReLU activation function
4. a FC layer with $7 \times 7 \times 128$ dimensionality
5. a Batch Normalization layer (128)
6. a ReLU activation Function
7. a Deconvolutional layer (a.k.a Convolutional Transpose) of dimension $(128 \times 64 \times 4 \times 2 \times 1)$
8. a BatchNormalization 2D layer (64)
9. a ReLU activation Function
10. a Deconvolutional layer of dimensions *image output*
11. a $\tanh(\cdot)$ function

Consequently, the backbone for $D(\cdot)$ is stated as follows:

1. a Convolutional layer of dimensions *image input* and classes representation (for CGAN and AC-GAN).
2. a leakyReLU activation function (0.2).
3. a Convolutional layer of dimensions $(64 \times 128 \times 4 \times 2 \times 1)$

¹⁶Available at: <http://yann.lecun.com/exdb/mnist/>

¹⁷Available at: <https://www.kaggle.com/zalando-research/fashionmnist>

¹⁸Available at: <https://www.cs.toronto.edu/~kriz/cifar.html>

¹⁹Available at: <https://www.cs.toronto.edu/~kriz/cifar.html>

Table 10: Experiments performed for each Dataset using the literature methods and the proposed GAF-GAN

Architecture	MNIST	Fashion-MNIST	CIFAR10	CIFAR100	COCO
GAF-GAN	X	X	X	X	X
GAF-ACGAN	X	X	X	X	X
GAF-WGAN	X	X	X	X	X
GAF-WGAN-GP	X	X	X	X	X
GAF-DRAGAN	X	X	X	X	X
GAF-LSGAN	X	X	X	X	X
GAF-CGAN	X	X	X	X	X

4. a Batch Normalization layer (128)
5. a LeakyReLU activation function
6. a FC layer with $7 \times 7 \times 128, 1024$ dimensionality
7. a Batch Normalization layer (1024)
8. a lakyReLU activation Function (0.2).
9. a FC Layer of fake/real and class dimension.

In all experiments, the ADAM optimization was employed with the same values for both $G(\cdot)$ and $D(\cdot)$. The parameters were: *Learning rate* = 0.0002, *beta1* = 0.5 and *beta2* = 0.999. The batch size is of 64 and the number of epochs was settled to *Epochs* = 100.

The experiments performed are outlined in table 10. From the entire set of GAN approaches, the ones described in this table were extended with the GAF approach.

The first experiment used the well known an extended MNIST Dataset. It consists of a low-resolution (28times28) images set of handwritten numbers from 0 to 9. Results are shown in figure 48. The experiment is shown only for the LS-GAN network. However, all results are shown in the annex of this thesis. MNIST dataset is not very complex, so networks learn the features of image numbers quickly. As it can be observed, there is not a significant difference between the original proposal and the GAF-GAN type.

The second experiment was run employing the Fashion-MNIST dataset. Results are drawn in figure 50. Similar to MNIST, there are not significant differences in the images obtained using both methods.

The third experiment was executed using CIFAR10 Dataset and it is shown in figure 52. As this dataset is more complex, including color channels, then the differences are appreciated more significantly. In fact, initial epochs were considered, due to the fact that GAF method impacts better in the early training stages, improving its performance. The left top images correspond to the original version of the ACGAN network, whereas the bottom presents exactly the same image in the same epoch 5, 25 and 50 respectively, for the GAF-ACGAN version. In fact, visual aspect of epoch 5 demonstrates the feasibility of GAF method. Furthermore, in epoch 25, visual inspection allows to establish that the object shapes are better defined. Finally, for epoch 50, the results are pretty similar as impact of the method is decreased. Similarly, the CIFAR100 dataset was also tested. The results are presented in

Figure 47: LS-GAN results on Fashion-MNIST Dataset for epochs 5, 25 and 50 respectively

Figure 48: GAF-LSGAN results on Fashion-MNIST Dataset for epochs 5, 25 and 50 respectively

figure 54. There is an enhancement in the images definition, although the improvements observed are not as significant as the CIFAR-10 case. The reason behind can be that the number of objects is large, and although the gossiped information can help the generator side, this improvement is not enough to allow it to learn the object's shape as fast as the CIFAR-10 dataset.

Finally, the COCO dataset was tested. Due to the high resolution of the images, compared to the previous datasets, a new architecture was proposed. An initial resize of the image to a $(224 \times 224 \times 3)$ size was applied. The architecture backbone for Generator is described as follows:

1. a Fully Connected layer FC of dimensions ($input - dim \times 512$)
2. a LeakyReLU activation function (0.2)
3. a Batch normalization (BatchNorm1d) (512)
4. a FC of dimensions (512, 1024)
5. a BatchNorm1d (1024)
6. a LeakyReLU (0.2)
7. a FC of dimensions $(1024, 1024 \times (256//64) \times (256//64))$
8. a BatchNorm1d $(1024 \times (256//64) \times (256//64))$
9. a LeakyReLU activation function (0.2)
10. a deconvolutional layer ConvTranspose2d $(1024 \times 512 \times 4 \times 2 \times 1)$
11. a BatchNorm2d (512)
12. a LeakyReLU activation (0.2)
13. a ConvTranspose2d of dimensions $(512 \times 256 \times 4 \times 2 \times 1)$
14. a BatchNorm2d (256)
15. a LeakyReLU activation (0.2)
16. a ConvTranspose2d of dimension $(256 \times 128 \times 4 \times 2 \times 1)$
17. a BatchNorm2d (128)
18. a LeakyReLU activation (0.2)
19. a ConvTranspose2d of dimensions $(128 \times 64 \times 4 \times 2 \times 1)$
20. a BatchNorm2d (64)
21. a LeakyReLU activation (0.2)
22. a ConvTranspose2d of dimensions $(64 \times 32 \times 4 \times 2 \times 1)$
23. a BatchNorm2d (32)
24. a LeakyReLU activation (0.2)

25. a ConvTranspose2d of dimensions $(32 \times 3 \times 4 \times 2 \times 1)$
26. a $\tanh(\cdot)$ function

Furthermore, in the Discriminating side, the architecture is stated as follows:

1. a Convolutional layer Conv2d of dimensions ($input - dim, out_channels = 32, kernel_size = 4, stride = 2, padding = 1$)
2. a LeakyReLU activation (0.2)
3. a Conv2d of dimensions ($in_channels = 32, out_channels = 64, kernel_size = 4, stride = 2, padding = 1$)
4. a BatchNorm2d (64)
5. a LeakyReLU activation (0.2)
6. a Conv2d of dimensions ($in_channels = 64, out_channels = 128, kernel_size = 4, stride = 2, padding = 1$)
7. a BatchNorm2d (128)
8. a LeakyReLU activation (0.2)
9. a Conv2d of dimension ($in_channels = 128, out_channels = 256, kernel_size = 4, stride = 2, padding = 1$)
10. a BatchNorm2d (256)
11. a LeakyReLU activation (0.2)
12. a Conv2d of dimensions ($in_channels = 256, out_channels = 512, kernel_size = 4, stride = 2, padding = 1$)
13. a BatchNorm2d (512)
14. a LeakyReLU activation (0.2)
15. a Conv2d ($in_channels = 512, out_channels = 1024, kernel_size = 4, stride = 2, padding = 1$)
16. a BatchNorm2d (1024)
17. a LeakyReLU activation (0.2)
18. a FC layer of dimensions $(1024 * (256//64) * (256//64), 1024)$
19. a BatchNorm1d (1024)
20. a LeakyReLU activation (0.2)
21. a FC connected layer of dimensions (1024, 512)
22. a BatchNorm1d (512)
23. a LeakyReLU (0.2)

Figure 49: LS-GAN results on Fashion-MNIST Dataset for epochs 5, 25 and 50 respectively

Figure 50: GAF-LSGAN results on Fashion-MNIST Dataset for epochs 5,25 and 50 respectively

24. a FC layer of $(512, output_{dim})$
25. a Sigmoid() activation function

The results are presented in figure 56. In this figure, it can be observed that the representation of the images created (*bottom*) using the GAF-GAN have a better appearance to the ones obtained by using the GAN by default. However, results look not much better for epoch 25 and 50 respectively.

It is important to highlight that the complexity of the COCO dataset does not allow to extract meaningful results in the initial epochs. Taking into account that the dataset contains more than 120 labeled objects and more than 330.000 images, it is hard to figure out the training to be near to distinguishable images.

The entire code, pre-trained models, dataset, parsers and examples presented here are available for the research community at:

<https://gitlab.com/tavitto16/phd-gaf.git>

Similarly, the examples for the rest of networks can be found in the repository

Figure 51: ACGAN results on CIFAR-10 for epochs 5, 25 and 50 respectively

Figure 52: GAF-ACGAN results on CIFAR-10 Dataset for epochs 5,25 and 50 respectively

Figure 53: ACGAN results on CIFAR-10 for epochs 5, 25 and 50 respectively

Figure 54: GAF-ACGAN results on CIFAR-100 Dataset for epochs 5,25 and 50 respectively

Figure 55: GAN results on COCO dataset for epochs 10, 33 and 55 respectively

Figure 56: GAF-GAN results on COCO Dataset for epochs 10,33 and 55 respectively

7 Chapter 7: Person Re-identification for Tracking Using Multi-modal Systems.

Chapter Summary

This section extends one of the main topics discussed: *Re-identification*. In this context, it is referred to associate the measurements gathered from diverse sensors to a single person. The idea is to collect data from the band as reference, and link it to the routes gathered by the Kinect, (also the Zenith if available) and the WSN. Over time, it involves several steps that are detailed here (synchronization, time-association and normalization). The work of this chapter is mainly related to the paper: “*Person Tracking using Multi-Modal Systems*”. 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) 2017 [41].

7.1 State-of-the-Art

Formally, the person identification problem arises when a target (object / person) must be tracked in an uniquely manner, in a particular scenario, and matched with a specific identity. This target can not be always monitored, however, it must be identified whenever possible. This problem is common in visual monitoring systems where the goal is to individualize targets that can be lost due to occlusions, lighting, etc. [155, 159]. In case the object is lost and detected again (by the same camera or a different one), the system must be able to determine that the new tracked element is the same to the one previously tracked.

The general approach to deal with this problem relies on the extraction of several features from the detected object. Some examples include RGB-based features such as color, shape, or texture [159, 150]; appearance features incorporate histogram, graph model, spatial occurrence model [155]; whereas biometric features comprise face patterns recognition and gait analysis [152, 151]. These features allow to describe a person and consequently match them with the one with the most similar detected elements.

This problem has attracted the focus of the scientific community due to its usefulness for a wide range of applications in which human activity recognition is required [158]. However, there exist several scenarios where visual features can not be directly employed due to privacy issues. Therefore, an alternative to camera-based systems is needed [153, 149, 156]. The use of depth sensors allows information analysis from infrared, skeletons, etc, which has opened a new scope in this research topic [151, 154].

Furthermore, due to the emergence of the use of wireless sensing technologies, human activity and health status can be easily monitored and analyzed [154, 156]. As a result, such systems can provide useful information in real-time not only to the user himself, but also to the user’s carer or doctor, since these technologies have been also adopted in health-related applications as well.

A widespread type of devices employed in such applications are sensor equipped Bracelets. The main use of these devices is for sports and activity tracking, and are usually equipped with IMU (Inertial Measurement Units) sensors that provide 3D acceleration and 3D angular velocity (gyroscope) information of the person’s movement, and calculate steps, distance, calories etc. However, some of

them are also equipped with health-related sensors that can acquire heart beat rate, body temperature and galvanic skin response.

The adoption of this technology in health tracking applications is very important and its contribution, especially for patients with mobility problems, is twofold. First, this technology can be employed by intelligent analysis systems that can evaluate the patient's movement and hence estimate the evolution of it. Thus, the medical professional, based on the movement analysis (time of the day, duration, frequency), can modify the patient's medication. Such analysis will be more accurate, objective and complete compared to systematic records kept by the patient. Second, since the system will be able to alert the patient's caregiver in case of emergency, patient autonomy will be extended.

In this work, a method for person individualization using skeleton information from Kinect v2 devices along with smart-wearables is described. This approach relies on the matching of movement features extracted from both devices. In subsection 7.2, the process of feature extraction as well as the main components of the proposed system are detailed. Subsection 7.3 shows the experiments performed and the obtained results.

Several works have been proposed in the literature to tackle the person re-identification problem using visual sensors [159], and extracting RGB-related features such as color, shape and texture [155].

Furthermore, proposals address this problem by using multiple cameras to extract the main visual features to be compared with the ones obtained by the other cameras [149]. Also, an important research effort has been dedicated to the extraction of features using visual depth sensors [150, 151].

Moreover, there exist approaches for re-identification using information from wireless devices such as smart-phones and wearables [153, 156]. Probably, the most similar work to the one presented here is [154]. In that paper, acceleration features are extracted from Bracelets and compared with the values obtained from depth cameras. However, in our method, orientation information is considered too, which can provide a more comprehensive representation of the performed actions, and therefore increase the accuracy of the identification process. Furthermore, the work presented is a particular case of the general re-identification problem, as the main aim is to individualize and identify persons wearing the bands and detected by the Kinect sensor.

The main contribution of this work is to provide a novel approach for dealing with the person individualization problem in monitoring systems, by using information from multiple sensors.

7.2 Proposed method

In this work, an indoor environment is considered, where a Kinect v2 sensor is deployed to detect and track people using only skeleton information extracted from depth images. As aforementioned, visual features are not employed, as the work presented is part of a project where privacy issues are critical.

The problem addressed in this work is the individualization and route extraction (tracking) of people in a room by associating measurements gathered from diverse data sources (Kinect depth and wearables). For this purpose, several assumptions have been made:

- People will be identified and tracked within the coverage range of depth sensor.
- Association will be performed for up to 6 skeletons (from 0 to 5) with the corresponding Bracelets

(from 1 to N).

- People not wearing Bracelets (i.e. detected only by Kinect) will not be associated. Analogously, data of people wearing Bracelets but not detected by Kinect will not be considered in the association process.
- It is assumed that a person is wearing a single Bracelet.
- All measurements are synchronized. A specific tool that gathers and synchronizes data has been developed for this purpose.

Let assume that there are P individuals in a room. Some of them wear Bracelets, thus there are $N < P$ Bracelets. Also, $D \leq 6$ skeletons are detected from Kinect, resulting in $2 \cdot D$ wrist joints to consider; we are not aware if the Bracelet is worn on the left or the right hand. Two approaches are proposed in this work for matching the signals from the two types of sensors. The first one matches the wrist Pitch extracted from both hands of the detected skeletons with the corresponding Pitch calculated from the Bracelets. The second method employs the Yaw information of the skeletons chest point and the estimated Yaw from the Bracelet sensors. In the next subsections, details regarding the calculation of these features (wrist Pitch and Yaw) for both sensors, data analysis and matching are presented.

7.2.1 Bracelet

It is assumed that the Bracelets employed can provide 3D acceleration and 3D angular velocity information. However, two important issues regarding these values must be taken into account:

- The coordinate system of the Bracelets is local, and therefore continuously changing. **(I1)**
- The gravity vector (which always points towards the earth's center) is part of the returned acceleration values **(I2)**. Namely, the acceleration values provided by the IMUs are equal to the vector sum of the acceleration due to the person's movement and the gravity force.

7.2.2 Kinect Sensor

In the case of the Kinect sensor, for each tracked person (up to 6 per Kinect device), the (x, y, z) coordinates for each of the 25 joints located on various parts of the detected human body²⁰ are provided. Initially, the skeleton joints are filtered using the Tobit-Kalman filter [147]. This is quite an important step, since the extracted skeleton often contains small erratic movements that significantly affect further calculations.

7.2.3 Calibration

According to **I1**, values extracted from the Bracelets and the Kinect skeletons can not be directly compared. As aforementioned, the coordinate system of the former is always changing according to the movement / orientation of the Bracelet sensor, while the latter uses a fixed coordinate system, which,

²⁰<https://msdn.microsoft.com/en-us/library/microsoft.kinect.jointtype.aspx>

however, also depends on the placement / Yaw of the sensor. As a result, the measured acceleration vectors differ between the Bracelet and the Kinect. Even trying to compare the magnitude of the acceleration values ($\sqrt{a_x^2 + a_y^2 + a_z^2}$), which is more robust since magnitude does not depend on the sensor orientation, would not be accurate due to **I2**. The reason is that the gravity force influences the Bracelet acceleration values. The direct removal (or addition) of the gravity force from the Bracelet (to Kinect) would not be a solution, since these processes will be performed in two different coordinate systems. Thus, in order to be able to compare the sensor outputs, a common coordinate system must be constructed as reference.

Bracelet calibration The Bracelet calibration method presented in [148] has been adopted, which estimates the orientation of the Bracelet by employing the information gathered by accelerometer and gyroscope. The output of this method is the orientation of the Bracelet device in relation to the starting point (initial location of the Bracelet). This information allows the elimination of gravity from acceleration measurements.

There are several parameters provided as a result of applying the calibration process. Some of these values involve the *declination* and the gyroscope β , which depend directly on the sensor technical specifications.

Kinect calibration Kinect has its own coordinate system. Its origin is located at the center of the IR sensor while X axis grows to the sensor's left, Y axis grows upwards and Z axis to the direction the sensor is facing. Thus, the Kinect coordinate system depends on the custom device setup. In order to calibrate the device, a Calibration tool has been created (Figure 57). In this software, the user selects a set of points (usually 6-10) and sets their coordinates in the desired coordinate system. Thus, it is possible to set the Z axis of the new coordinate system to be perpendicular to the ground, and the XY plane to identify with the floor plane. The Calibration tool can retrieve the coordinates of the selected points on the Kinect coordinate system and then calculate a transformation between these two coordinate systems. In this tool, the coordinates in both systems are given in centimeters, thus the required transformation is a translation and a rotation (no scaling is required). For the rotation and translation matrices calculation, the Singular Value Decomposition (SVD) of the covariance matrix of the coordinates of the two systems is employed.

7.2.4 Feature calculation

In this section, the aforementioned features will be calculated from both Bracelet and Kinect devices. The following notation will be used for the vectors: $\mathbf{m}_{d,a}^c$ where $\mathbf{m} \in \{p, v, a, g\}$ denotes the measurement (p : position, v : velocity, a : acceleration, g : angular acceleration), $d \in \{b, k(j)\}$ the device (b : Bracelet, k : Kinect, j : joints index), $a \in \{x, y, z\}$ the axis and c denotes that the vector is on the calibrated space. Hence, the Bracelet provides the acceleration $\mathbf{a}_{b,x}, \mathbf{a}_{b,y}, \mathbf{a}_{b,z}$ and gyroscope $\mathbf{g}_{b,x}, \mathbf{g}_{b,y}, \mathbf{g}_{b,z}$ vectors from which the bracelet orientation is calculated (given in quaternions $Q_i, i = 1, 2, 3, 4$) that is used to extract the linear acceleration²¹ $\mathbf{a}_{b,x}^c, \mathbf{a}_{b,y}^c, \mathbf{a}_{b,z}^c$ in the calibrated space. Similarly, Kinect provides the coordinates of the skeleton wrists and elbows that after the calibration result in $\mathbf{p}_{k(i),x}^c, \mathbf{p}_{k(i),y}^c, \mathbf{p}_{k(i),z}^c$ where $i \in \{5 \text{ (left elbow)}, 6 \text{ (left wrist)}, 9 \text{ (right elbow)}, 10 \text{ (right wrist)}\}$.

²¹The actual device acceleration excluding the gravity force.

Figure 57: Kinect calibration tool

Figure 58: Pitch calculation

Figure 59: Bipartite graph representing the association links of the N Bracelets with the available skeletons

Wrist pitch

In Kinect case, the wrist pitch can be calculated from the wrist and elbow coordinates. Assuming (Figure 58) that the elbow is the point O and the wrist the point A , then the pitch is the angle θ equals $\tan^{-1}(AB/OB)$. Since $OB = \sqrt{CB^2 + BD^2}$, $\theta = \tan^{-1}(AB/\sqrt{CB^2 + BD^2})$. Taking into account that $AB = \mathbf{p}_{k(6|10),z}^c - \mathbf{p}_{k(5|9),z}^c$, $BC = \mathbf{p}_{k(6|10),y}^c - \mathbf{p}_{k(5|9),y}^c$ and $BD = \mathbf{p}_{k(6|10),x}^c - \mathbf{p}_{k(5|9),x}^c$, then the wrist pitch $\theta_{l|r}$ for the left and right wrist is given by:

$$\theta_{l|r} = \tan^{-1} \left(\frac{d_{l|r,z}}{\sqrt{d_{l|r,x}^2 + d_{l|r,y}^2}} \right) \quad (77)$$

where d is the distance between the wrist and elbow for each axis, namely $d_{l,a} = \mathbf{p}_{k(6),a}^c - \mathbf{p}_{k(5),a}^c$ and $d_{r,a} = \mathbf{p}_{k(11),a}^c - \mathbf{p}_{k(10),a}^c$, $a \in \{x, y, z\}$.

The Bracelet based wrist pitch estimation can be performed by employing the Bracelet orientation. By using the orientation quaternions and following the same procedure with the Kinect case, it is possible to apply a similar equation with (77) namely

$$\theta = \tan^{-1} \left(\frac{Q_3}{\sqrt{Q_1^2 + Q_2^2}} \right) \quad (78)$$

where Q_i are the quaternion elements of the Bracelet orientation.

Figure 60: Example of Yaw variation over time for both sensors. This case corresponds to the episode 3 described in section 7.3. On the left, the MSBand (blue) matches with the cyan skeleton. On the right hand, the other MSBand in the experiment matches with the red Skeleton. Notice that the green and cyan lines correspond to the same person whose skeleton was lost and recovered again.

Orientation extraction and comparison

As previously mentioned, tools implemented in [148] allow to extract the Yaw, Pitch and Roll features from IMU sensors. According to Madgwick filter, the Yaw is calculated as follows:

$$Yaw = \arctan \left(\frac{2(Q_2 \cdot Q_3 + Q_1 \cdot Q_4)}{Q_1^2 + Q_2^2 - Q_3^2 - Q_4^2} \right), \quad (79)$$

where Q denotes the quaternion parameters as previously described in section 7.2.3. The outcome of this filter is the route in terms of orientation (Yaw), which indicates how much the person has turned around over the detected time.

However, Yaw values are relative to the initial orientation of the device; in this case the Bracelets. Therefore, the result of this filter is the angle between the current and the initial position around the yaw (z) axis, and not between the current position and the magnetic north.

For the Kinect orientation calculation, a central skeleton joint is selected (i.e. the 20-th that corresponds to the chest). The orientation of the line segment defined by the consecutive chest joint coordinates $(P_{k(20),x}^{c,t}, P_{k(20),y}^{c,t})$ and $(P_{k(20),x}^{c,t+1}, P_{k(20),y}^{c,t+1})$ at times t and $t + 1$ respectively, corresponds to the Yaw of the person (Figure 58) in the Kinect fixed coordinate system.

As a consequence, the association process consists in searching for similar orientation routes among the Bracelets and Kinect skeletons.

Let $Y_b(t)$ and $Y_k(t)$ be the extracted Yaw sequences from Bracelet and Kinect device respectively. If both are extract from the same person, we expect $Y_b(t) \simeq Y_k(t) + s$, where s represents the angle difference between the two coordinate systems (Kinect coordinate system and Bracelet initial Yaw).

In order to eliminate s , instead of comparing the extracted Yaw sequences, we compare their first order derivatives, namely the $dy_b(t)/dt$ and $dy_k(t)/dt$. As a result, if both Yaw sequences are extracted from the same person, $dy_b(t)/dt \simeq dy_k(t)/dt$.

Data association Process The association process is performed using the Hungarian algorithm [161]. It computes a complete matching of the bipartite graph (Figure 59), such that the total error of the matched elements is minimized.

The link weights are a combination of the MSEs obtained for every feature. Therefore:

$$c_{ij} = \sqrt{e_{pitch}(ij) + e_{Yaw}(ij)} \quad (80)$$

The cost for each link is given by the Mean Square Error (MSE) for both pitch and orientation. As a result, the combination of Bracelets with the skeletons is obtained. A modification of the standard algorithm was performed, in order to avoid multiple choices for a particular skeleton.

The proposed modified Hungarian algorithm contains the following two additional rules:

Theorem 1 *If in any step a Bracelet has the minimum cost for more than one Kinect skeleton, the algorithm will associate the one with the lesser error and remove it. Similarly, in case of having a Kinect skeleton with a minimum cost for several bands, the algorithm will chose the one with the lesser error and will remove it from the list.*

Theorem 2 *If the error values are above a certain threshold, the Bracelet and skeleton will not be associated.*

The former condition is put in place to avoid multiple selections of a single skeleton for a Bracelet candidate. The latter, aims to avoid associating a skeleton with a Bracelet, even if algorithm yields this result, because the error is large. The threshold is empirically obtained and depends on the time-window chosen. For the experiments considered in this work, the time-window was 1 minute and the threshold was set to 2500.

The steps for the implementation of the proposed algorithm are described as follows:

Calibration process of sensors

- 1: Calibration of Bracelets as shown in 7.2.3
- 2: Calibration of Kinect Sensor as shown in 7.2.3

Feature Extraction

- 3: Orientation calculation from Kinect as presented in section 7.2.4
 - 4: Orientation calculation from Bracelets as described in section 7.2.4
 - 5: Association of Bracelet-skeleton as described in section 7.2.4
-

7.3 Experimental results

In this section, the results of the experiments with the proposed algorithms are detailed. A commercial band developed by Microsoft was employed as wrist device²². According to the technical specifications, this band is able to provide 8, 31 or 62 measurements per second. This device is denoted as *MSband*.

Furthermore, a Kinect sensor was deployed in the monitored room at a height of 250cm. The experiment was performed under two lighting conditions in order to force the Kinect device into both available frame rates (15 and 30 fps).

During the experiments performed, any time difference between the coupled samples of the two sensors was detected. However, without synchronization, differences between timestamps varied up to 500ms. Additionally, the MSE obtained from the synchronized and non-synchronized data comparison reaches up to a 10% larger which seriously affects the performance of the proposed algorithm.

Regarding device synchronization, differences between the unsynchronized sensor timestamps varied up to 500ms and the obtained MSE from the synchronized and unsynchronized data were up to a 10% higher. These observations demonstrate that device synchronization is essential for accurate results.

To test the feasibility of the proposed algorithms, the following episodes were defined:

- *Epi1*: A person is detected by the Kinect sensor (skeleton), the person walks for a while and afterwards leaves the room.
- *Epi2*: A person is detected by the Kinect sensor (skeleton), suddenly the tracker loses the person, and after a while it re-detects the person.
- *Epi3*: Two people are detected by the Kinect sensor (skeleton), and both wear Bracelets. They walk for a while fully detected by Kinect, and then they go out of the scene.
- *Epi4*: Two individuals are detected by the Kinect sensor (skeleton), but only one of them is wearing a Bracelet.
- *Epi5*: Two people are detected by the Kinect sensor (skeleton), both wearing Bracelets. Suddenly, one of them is lost by Kinect and after a while the person is re-detected.
- *Epi6*: Three persons are detected by the Kinect sensor (skeleton), with two of them wearing Bracelets. Kinect loses and re-detects them as they walk in front of each other.

The Kinect sensor was calibrated so that the *XY* plane matches the room floor. Sessions of diverse durations have been recorded in which (2-3) people were present on each one. In Table 11 information about the recorded sessions is presented. More specifically, the episode id, the duration in seconds and the number of people / unique skeleton tracks that Kinect was able to detect are shown. In some sessions, the number of skeleton tracks is higher than the number of people, since some people have been re-detected due to occlusions. Finally, the last two columns on the right 11 contain the successful detections for the two methods. It can be observed that both of these methods achieve a high matching accuracy.

²²<https://support.microsoft.com/en-us/help/4000325/band-hardware-specifications>

Table 11: Information about the recorded sessions (in parenthesis the number of people wearing Bracelet)

id	Duration	# of people / skeletons	Pitch	Yaw
1	60	1 (1) / 1	1	1
2	60	2 (0) / 3	3	3
3	45	2 (2) / 4	3	4
4	35	2 (1) / 3	2	3
5	90	3 (2) / 7	6	6
6	35	3 (2) / 5	1	3

An example matching using the (Yaw) orientation for two individuals is shown in Figure 60. The vertical axis represents the Yaw and the horizontal axis the time. This particular graphic represents the *Epi3* recording. On the left side, after the green skeleton is lost and re-detected as cyan, it can be seen that the Bracelet time-series (blue) follows closely the cyan skeleton, which is the correct match. Similarly, on the right side, the time-series of the other Bracelet (blue) closely follows the red skeleton, which is again the correct match.

The entire dataset with the information of sensors, calibration and results shown in this work is available at <http://www.gatv.ssr.upm.es/~abh/>

8 Chapter 8: Conclusions and Future Work.

Across this doctoral thesis, a range of algorithms were proposed to cope with the ambitions presented in the objectives chapter. A research methodology has been defined and followed in the project (and thesis) timeline according to the expertise of work context (multi-disciplinary research involving healthcare professionals and technicians).

It is important to highlight that the algorithms and techniques developed do not form a system itself. Although, as it was observed, these are the basis of complex analytic systems that aim at providing a holistic solution to one of the key problems in our society. In some words, the algorithms presented here can be used to improve/release innovative solutions for areas where human behavior knowledge can provide add-value. As mentioned in this thesis, healthcare was the main focus of the value proposition, however, it is not the only usability of these algorithms, but any application for indoor positioning and tracking.

For the development of these techniques, essentially, three main topics (developed in the 4 main thesis chapters), where the contributions can be outlined as follows:

- There was a consistent work in the development of efficient systems able to collect relevant information from users to be processed and extract meaningful patterns for the movement analysis. This information can be directly employed for behavior detection (chapter 4) or be used as input of more complex systems. The main outcomes of this topic are in the definition, collection, processing and extraction of meaningful features for high level algorithms inference.
- In the context of indoor detection, there are several contributions in this work:
 1. The description and development of integrated systems for indoor positioning using commercial technologies. The use of *fingerprinting* technique to create a ground-truth and characterize the scenario under monitoring in combination with multiple manners of position estimation. The work was focused on the development of methods for Position estimation (referred as TPE in this thesis), and tracking using known techniques based on Kalman Filter. This work was defined as the MUFAF algorithm.
 2. The second proposition was centered on the use of Deep Learning techniques for improving the in-cell localization results. As described in chapter 5. The use a Neural network to characterize the in-cell positioning provides this type of system with a larger resolution, improving the accuracy results
- The proposition of an algorithm for person re-identification in indoor environments based on the data collected from multiple sensors. This approach
- The use of Deep Learning, in concrete Generative Adversarial Networks GAN, was proposed to increase the amount of data available in the fingerprinting database to significantly improve the results obtained. It was shown that the estimation error can be reduced by using this approach.
- Finally, a novel architecture to speed the training process in images generation was proposed. One of the main challenges in Generative Adversarial models is to find a good configuration to reach convergence. The proposal of using a reduced part of the feature maps in the training phase speeds this procedure. As a consequence, the variety of the features learned is reduced, however, a weight to control the learning maps inclusion is proposed.

From a technical perspective, the main limitation of the presented system is given by the search of the best trade-off between coverage and accuracy of the sensors considered. Kinect sensor has large precision, however, the maximum distance is around 4m and 120 degrees view. Furthermore, WSN detection has a large coverage (up to 10m bluetooth and 40m Xbee). Nonetheless, accuracy of estimations is lower than Kinect and Zenith camera as shown in Table 4. Finally, Zenith cameras have a wide coverage range (up to 20m), however, occlusions and angle distortion decrease the tracking performance compared to the accuracy obtained with the Kinect sensor.

Finally, conversely to tailored solutions for healthcare monitoring which are generally expensive, most of the devices composing the system presented are well known and of general purpose. The Kinect sensor is available to be acquired for around 150USD. Moreover, Zenith camera employed in this work is around 250USD, and the bands cost 100USD [20] and 250USD [18] respectively. There is a wide range of WSN technologies and the prices are around 100USD. Finally, the Kinect sensor requires at least 1 USB 3.0 port and 8GB RAM to work.

Conclusions: Chapter 3

Initially, a complete multi-modal scheme for elder people monitoring was presented. The system architecture was outlined as well as the main modules such as DAP, ABD of the low level subsystem. The key concept presented in this work is the multi-modal approach, which includes a set of sensors deployed to detect abnormal events in scenarios associated to PD patients. It has been shown that accuracy of the estimations for the events considered is high as well as the usefulness of this approach for multiple environments (e.g. home, rehabilitation therapy). Results confirm the modularity of the proposed system as the system performance was good even when not all sensors were available. As a conclusion, it has been demonstrated that the multi-modal approach allows to improve the accuracy of locations estimated which is of special interest for the PDs patients' monitoring. Additionally, WSN in combination with the trajectories can be used to detect anomalies such as the patients leaving the house.

Moreover, the fusion of bracelet sensors and skeleton body points allows to perform an activity analysis. The individualization process described in this work endows the system with a multi-user functionality which is interested for places with concurrent patients such as daily centers. ABD module permits to identify unexpected behaviors that might require immediate attention. The time-line evolution of these events will allow professionals to better understand the progress of the illness and make the proper decisions.

Conclusions: Chapter 4

In the previous chapter, the focus was pointed to the non-invasive methods for indoor tracking topic. This topic is of particular interest to healthcare applications due to the possibility of assess the behavior of persons and extract meaningful patterns. In concrete, several approaches were presented. From the technological point of view, multiple sensors were chosen to perform experiments of how the proper combination of the signals received by Access Points can provide estimations of the patient/user position and, how by means of extended techniques such as the Kalman Filter, the quality of the estimations can be improved.

In this approach, firstly a fingerprinting method was detailed. This method allows to characterise the space under monitoring by splitting the area into sub-cells of a predefined size. In each cell, information of propagation was collected for a determine time period. Two architectures were presented to cope with indoor localization based on the information available. The advantages and drawbacks of each proposal were analyzed and the results shown that the average accuracy is large yielding to errors lesser than 1m, which is reasonable for such systems.

Moreover, this approach can be further improved, by appending some extra component. In addition to the position estimation, the identification presents a significant breakthrough of these systems. Two approaches that identify a person in an indoor environment using movement features from multiple sensors have been presented. The methods rely on the extraction and association of Pitch and Yaw orientation from Bracelets and Kinect. Experimental results have shown the validity of both approaches. As a future work, it will be interesting to evaluate the scalability of the system in crowded and large-scale scenarios, as well as to explore the inclusion of sensors such as Wireless Sensor Networks (WSN) tracking systems, which can significantly increase the accuracy and range of the presented algorithms.

Conclusions: Chapter 5

In this chapter, a complete overview on Fingerprint construction and assessment has been presented. On the one hand, this paper shows the impact of different positioning methods using RSSI Fingerprinting technique on the estimation accuracy. On the other hand, a custom filter is proposed to improve the positioning results and finally, a Tracking Algorithm is applied to adjust the measurements to a known movement model.

Moreover, the Fingerprint dataset was constructed taken into account the Fingerprint parameters for diverse situations. Important aspects such as the orientation and human presence were considered and it has been shown that a significant improvement is attained compared to other estimation procedures. These parameters allow to obtain a consistent feature vector for training Machine/Deep Learning algorithms. The inclusion of the Yaw/Heading parameter improves the accuracy up to 15% compared to the results obtained without this parameter and outperforming the common deterministic and statistical methods up to 45 – 50%.

The proposed Gaussian Outliers Filter with the weighted combination of the output probabilities yields to an increase in the accuracy of a 10 – 15% avoiding impossible states or low probable estimations.

Deep Learning approaches provide better results than Machine Learning algorithms when the input data is well conditioned and representative. In this work, these methods present better results than Machine Learning techniques. More complex patterns can be learned by the Neural Networks that fit better over the input data.

Particle Filter is finally applied to adjust the final estimations to a realistic movement model. The use of the estimated velocity and the Yaw/Heading angles outperform the final accuracy presented by all the methods without applying a Tracking Algorithm. An improvement of 25% is achieved after using this filter over the Fingerprinting estimations.

Conclusions: Chapter 6

In this chapter, a novel framework for performing *Data Augmentation* was proposed. The model relies on the use of Generative Adversarial Networks to increase the number of samples in the fingerprinting model. Two network architectures were presented: (a) The Conditional Generative Adversarial Network and (b) The Auxiliary Conditional GAN were employed. As it could be observed in the results, there is a significant estimation error reducing. The network was trained to identify the propagation pattern in each of the Fingerprinting cells.

Moreover, in this chapter, it was proposed a novel architecture framework to speed up the training process in GAN networks. The idea is to use a weighted version of the training feature-maps to “gossip” the training characteristics to the corresponding layer in the generative side. The proposal relies in the assumption of mirror backbones for the generative side and the discriminator. The process starts in the classification side, where the feature maps among the diverse layers are stored. Then these feature maps are weighted to be added to the generative side. The goal is on to speed the training of the generative module.

Results in the MNIST dataset were acceptable although not make a big difference with respect to the GAN architecture.

This approach reached a significant improvement with respect to all GAN architectures presented in the literature for CIFAR-10, CIFAR-100 y COCO Datasets.

Conclusions: Chapter 7

In this chapter, two approaches that identify a person in an indoor environment using movement features from multiple sensors have been presented. The methods rely on the extraction and association of Pitch and Yaw orientation from Bracelets and Kinect. Experimental results have shown the validity of both approaches. As a future work, it will be interesting to evaluate the scalability of the system in crowded and large-scale scenarios, as well as to explore the inclusion of sensors such as Wireless Sensor Networks (WSN) tracking systems, which can significantly increase the accuracy and range of the presented algorithms.

8.1 Future Lines and Dissemination results

The work done in this thesis is still subject to improvement, and it is just part of a European Project that will be further developed in the upcoming years. From a general perspective, it is important to highlight that most of the systems are being integrated under standards such HL7 or FHIR that guarantee interoperability. However, there still exists a huge amount of issues and open questions in terms of data: data privacy, data ownership, security, and integrity.

From a technical perspective, across this thesis, it can be observed the evolution that most of the systems are having into Deep Learning approaches. In fact, the initial chapters have a strong component of statistical methods to reach the positioning and tracking of individuals, whereas the latter chapters mainly rely on the use of Neural Networks for the estimation and Generative models for data augmentation.

Therefore, a clear line of investigation is given by going into further replacement of parts of the systems and workflow for person estimation by Neural Network approaches. Fingerprinting is a valid technique that will remain, and the analysis presented in this thesis allow to increase this database to gather multiple scenarios. However, the analysis and the fingerprinting varies in a day as shown in chapter 6. Therefore, there is still an open opportunity to create models able to adapt not only to the scenario conditions in terms of propagation, but also in time (and unexpected changes).

Moreover, the investigation in the field of Generative Models is under continuous development. By the time of presenting this work, still simulations are running to improve in more complex datasets, being able to generate images adapted to the dataset conditions. As an example, COCO dataset is labeled with multiple information that can be embedded to be inserted in the Generative Network. In fact, even the parameter that controls the weight of the feature maps gossiped can be change deceptively according to the loss function.

Regarding the technique for re-identification, it is important to highlight that this is a hot topic that attracts a lot of research efforts. Currently, this work is being extended by proposals that face with multiple features, and that extract a lot of indicators that overcome the results presented here.

From practical aspects that shall be checked, a study of the impact of varying the number of nodes in the network can be performed. Moreover, a deep analysis of techniques for Fingerprint auto-recalibration of over time can yield fast and adaptive systems that respond to the indoor scenario changes. Novel techniques such as Reinforcement Learning can be applied to generate new samples using the previous collected data. Additionally, the research focus of the data fusion of multiple modalities from WSN with visual and depth sensors is relevant to overcome the problems associated to these sensors. The appropriate fusion can solve several tracking problems with cameras such as occlusions, lighting conditions and coverage among others.

Additionally, several international papers related to the subjects mentioned in this work were reached. Further, some relevant ideas in the thesis lifetime were proposed as basis of implementations, test-beds, or boosting future projects and work-lines:

- A. Belmonte-Hernández, Hernández-Peñaloza, Gustavo, “*Person tracking association using multi-modal systems*,” 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Lecce, 2017, pp. 1-6 [41].
- Belmonte Hernández, Alberto ; Daras, Petros ; Theodoridis, Thomas ; Burgos González, Marta ; Hernández-Peñaloza, Gustavo ; Solachidis, Vassilios ; Jiménez Ramos, Jennifer ; Alvarez, Federico ; Vretos, Nicholas ; Carrasco, Laura. “*A novel framework for physical therapy rehabilitation monitoring and assessment in Parkinson disease patients using depth information*”, in Proceedings of the 12th ACM International Conference on Pervasive Technologies Related to Assistive Environments. 535-539. 10.1145/3316782.3322759 (2019). [43].
- Lopez Velasco, Juan ; D. Martín, Velasco ; Moreno, Francisco ; Hernández-Peñaloza, Gustavo ; Burgos, M ; Alvarez, Federico ; Marín, M ; Carrasco, L. (2018). “*Acceptance of Cognitive Games through Smart TV Applications in Patients with Parkinson’s Disease*”. Proceedings of the 11th Pervasive Technologies Related to Assistive Environments Conference 428-433. 10.1145/3197768.3201553 [42].

As future work, several activities can be integrated to this approach. As an example, an analysis of medical information can be employed for the implementation of algorithms for multi-modal Fusion and

Thesis Contents	Publication
3	F. Alvarez et al., "Behavior Analysis through Multimodal Sensing for Care of Parkinson's and Alzheimer's Patients," in IEEE MultiMedia, vol. 25, no. 1, pp. 14-25, Jan.-Mar. 2018. and G. Hernández-Peñaloza, A. Belmonte-Hernández, M. Quintana and F. Álvarez, "A Multi-Sensor Fusion Scheme to Increase Life Autonomy of Elderly People With Cognitive Problems," in IEEE Access, vol. 6, pp. 12775-12789, 2018.
4	G. Hernández-Peñaloza, A. Belmonte-Hernández, M. Quintana and F. Álvarez, "A Multi-Sensor Fusion Scheme to Increase Life Autonomy of Elderly People With Cognitive Problems," in IEEE Access, vol. 6, pp. 12775-12789, 2018.
5	A. Belmonte-Hernández, G. Hernández-Peñaloza, D. Martín Gutiérrez and F. Álvarez, "SWiBluX: Multi-Sensor Deep Learning Fingerprint for Precise Real-Time Indoor Tracking," in IEEE Sensors Journal, vol. 19, no. 9, pp. 3473-3486, 1 May1, 2019.
6	G. Hernández-Peñaloza, A. Belmonte-Hernández, David Martín , F. Álvarez. "Recurrent model for wireless indoor tracking and positioning recovering using generative networks," Submitted to IEEE Sensors Journal, Aug 2019.
7	A. Belmonte-Hernández, G. Hernández-Peñaloza, et al. "Person Tracking using Multi-Modal Systems" 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Lecce, 2017.

Table 12: Scientific Q1 Ranked Journals published during this thesis lifetime.

analysis. Additionally, action points to fuse physical activities with subjective patients information can increase the data available for the creation of inferences in the Clinical Decision Support System (CDS).

References

- [1] Alzheimer's Disease International, "*Dementia and Risk Reduction An analysis of protective and modifiable factors*," World Alzheimer Report 2014.
- [2] Kulisevsky J, Pagonabarraga J. "*Cognitive impairment in Parkinson's disease: tools for diagnosis and assessment*," Mov Disord. 2009.
- [3] M. Sajjad; S. Khan; Z. Jan; K. Muhammad; H. Moon; J. T. Kwak; S. Rho; S. W. Baik; I. Mehmood, "*Leukocytes Classification and Segmentation in Microscopic Blood Smear: A Resource-Aware Healthcare Service in Smart Cities*," in IEEE Access, 2016.
- [4] H. Medjahed, D. Istrate, J. Boudy, J. L. Baldinger and B. Dorizzi, "*A pervasive multi-sensor data fusion for smart home healthcare monitoring*," 2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011), Taipei, 2011.
- [5] A. Alaiad and L. Zhou, "*Patients' Adoption of WSN-Based Smart Home Healthcare Systems: An Integrated Model of Facilitators and Barriers*," in IEEE Transactions on Professional Communication, 2017.
- [6] A. Mathur, T. Newe and M. Rao, "*Healthcare WSN: Cluster Elections and Selective Forwarding Defense*," 2015 9th International Conference on Next Generation Mobile Applications, Services and Technologies, Cambridge, 2015.
- [7] R. Lun, C. Gordon and W. Zhao, "*Tracking the activities of daily lives: An integrated approach*," 2016 Future Technologies Conference (FTC), San Francisco, CA, 2016.
- [8] Vimarlund V, Olve NG, Scandurra I, Koch S. "*Organizational effects of information and communication technology (ICT) in elderly homecare: a case study*," Health Informatics J.
- [9] K. Rybova and J. Slavik, "*Smart cities and ageing population - Implications for waste management in the Czech Republic*," 2016 Smart Cities Symposium Prague (SCSP), Prague, 2016.
- [10] Kamel Boulos MN, Al-Shorbaji NM. "*On the Internet of Things, smart cities and the WHO Healthy Cities*," Int J Health Geogr. 2014.
- [11] Moreno FÁ, Merchán-Baeza JA2, González-Sánchez M3,4, González-Jiménez J5, Cuesta-Vargas AI6,7. "*Experimental Validation of Depth Cameras for the Parameterization of Functional Balance of Patients in Clinical Tests*," Sensors (Basel) 2017.
- [12] Technical specifications: <http://www.vivotek.com/fe8180/#specifications>
- [13] Santana, Zambom E, Gerosa A, Kon M, Milojevic F, Milojevic D. "*Software platforms for smart cities: Concepts, requirements, challenges, and a unified reference architecture*," arXiv preprint arXiv:1609.08089, 2016.
- [14] G. Merlino, D. Bruneo, F. Longo, A. Puliafito and S. Distefano, "*Software Defined Cities: A Novel Paradigm for Smart Cities through IoT Clouds*," 2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom), Beijing, 2016

- [15] C. Doukas and F. Antonelli, "A full end-to-end platform as a service for smart city applications," 2014 IEEE 10th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), Larnaca, 2014.
- [16] A. Costanzo, A. Faro, D. Giordano and C. Pino, "Mobile cyber physical systems for health care: Functions, ambient ontology and e-diagnostics," 2016 13th IEEE Annual Consumer Communications Networking Conference (CCNC), Las Vegas, NV, 2016.
- [17] Y. Zhang, F. Wang and S. Yu, "Use Fukunaga-Koontz Transform to Solve Occlusion Problems in Multitarget Tracking," 2009 Chinese Conference on Pattern Recognition, Nanjing, 2009.
- [18] Microsoft Band specifications: <https://support.microsoft.com/en-us/help/4000325/band-hardware-specifications>
- [19] Kinect Sensor specifications: <https://developer.microsoft.com/en-us/windows/kinect/hardware>
- [20] Hexiwear specifications. <http://www.hexiwear.com/hardware/>
- [21] IEEE802.15.4 Wireless Personal Area Networks WPAN Task group 4, <http://www.ieee802.org/15/pub/TG4.html>
- [22] Kinect SDK, <https://developer.microsoft.com/es-es/windows/kinect>
- [23] S. Yiu, M. Dashti, H. Claussen, F. Perez-Cruz, "Locating user equipments and access points using RSSI fingerprints: A Gaussian process approach," 2016 IEEE International Conference on Communications (ICC), Kuala Lumpur, 2016.
- [24] J. Y. Zhu, A. X. Zheng, J. Xu, V. O. K. Li, "Spatio-temporal (S-T) similarity model for constructing WIFI-based RSSI fingerprinting map for indoor localization," Indoor Positioning and Indoor Navigation (IPIN), 2014 International Conference on, Busan, 2014.
- [25] S. Doğan, F. Yılmaz, S. Ertürk, "On the error performance bounds of positioning using RSSI-based fingerprints," 2015 23rd Signal Processing and Communications Applications Conference (SIU), Malatya, 2015.
- [26] JooSeuk Kim, C. Scott, "Robust kernel density estimation," 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV, 2008.
- [27] V. Akbarzadeh, C. Gagné, M. Parizeau, "Kernel density estimation for target trajectory prediction," Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on, Hamburg, 2015.
- [28] A. Kushki, K.N. Plataniotis, A.N. Venetsanopoulol, C. S. Regazzoni, "Radio Map Fusion for Indoor Positioning in Wireless Local Area Networks," 2005 7th International Conference on Information Fusion (FUSION).
- [29] Ho N-H, Truong PH, Jeong G-M, "Step-Detection and Adaptive Step-Length Estimation for Pedestrian Dead-Reckoning at Various Walking Speeds Using a Smartphone," Sabatini AM, ed. Sensors (Basel, Switzerland) 2016.
- [30] Neeti Wagle, Eric Frew. "A Particle Filter Approach to WiFi Target Localization," AIAA Guidance, Navigation, and Control Conference. Toronto, Ontario, Canada.

-
- [31] Felix Scholkmann, Jens Boss, Martin Wolf. “*An Efficient Algorithm for Automatic Peak Detection in Noisy Periodic and Quasi-Periodic Signals*,” Algorithms 2012.
 - [32] B. Li, C. Yang, Q. Zhang and G. Xu, “*Condensation-based multi-person detection and tracking with HOG and LBP*,” 2014 IEEE International Conference on Information and Automation (ICIA), Hailar, 2014.
 - [33] Scaramuzza, D., Martinelli, A. and Siegwart, “*A Toolbox for Easy Calibrating Omnidirectional Cameras*,” Proceedings to IEEE International Conference on Intelligent Robots and Systems (IROS 2006), Beijing China, 2006.
 - [34] J. Gui, T. Zhang and L. Liu, “*The Application of Kalman Filtering in the System of Image Difference Moving Objects Tracking*,” 2011 International Conference on Computational and Information Sciences, Chengdu, China, 2011.
 - [35] T. D. Gottschalk, “*Concurrent Implementation of Munkres Algorithm*,” Proceedings of the Fifth Distributed Memory Computing Conference, 1990.
 - [36] A. Lorusso, D. W. Eggert, R. B. Fisher, “*A Comparison of Four Algorithms for Estimating 3-D Rigid Transformations*,” British Machine Vision Conference.
 - [37] G. Hernández-Peñaloza, A Belmonte-Hernández, M Quintana, F. Álvarez. “*A Multi-Sensor Fusion Scheme to Increase Life Autonomy of Elderly People With Cognitive Problems*,” in IEEE Access, vol. 6, pp. 12775-12789, 2018.
 - [38] A Belmonte-Hernández, G. Hernández-Peñaloza, F. Álvarez, G Conti “*Adaptive Fingerprinting in Multi-Sensor Fusion for Accurate Indoor Tracking*,” in IEEE Sensors Journal, vol. 17, no. 15, pp. 4983-4998, Aug.1, 1 2017.
 - [39] A Belmonte-Hernández, G. Hernández-Peñaloza, David Martín , F. Álvarez. “*SWiBluX: Multi-Sensor Deep Learning Fingerprint for Precise Real-Time Indoor Tracking*,” in IEEE Sensors Journal, vol. 19, no. 9, pp. 3473-3486, 1 May1, 2019.
 - [40] G. Hernández-Peñaloza, A Belmonte-Hernández, David Martín , F. Álvarez. “*Recurrent model for wireless indoor tracking and positioning recovering using generative networks*,” Submitted to IEEE Sensors Journal, Aug 2019.
 - [41] A Belmonte-Hernández, G. Hernández-Peñaloza, et al. “*Person Tracking using Multi-Modal Systems*”. 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) 2017.
 - [42] López Velasco, Juan, et al. “*Acceptance of Cognitive Games through Smart TV Applications in Patients with Parkinson’s Disease*”, in Proceedings of the 11th Pervasive Technologies Related to Assistive Environments Conference 2018.
 - [43] Belmonte Hernández, Alberto ; Daras, Petros ; Theodoridis, Thomas ; Burgos González, Marta ; Hernández-Peñaloza, Gustavo ; Solachidis, Vassilios ; Jiménez Ramos, Jennifer ; Alvarez, Federico ; Vretos, Nicholas ; Carrasco, Laura. “*A novel framework for physical therapy rehabilitation monitoring and assessment in Parkinson disease patients using depth information*”. Proceedings of the 12th ACM International Conference on Pervasive Technologies Related to Assistive Environments 2019.

-
- [44] S. Yokota, D. Chugo, H. Hashimoto and K. Kawabata, “*Visual impression to robot motion imitating human - study on delay motion*,” 2016 9th International Conference on Human System Interactions (HSI), Portsmouth, 2016.
 - [45] John G. Proakis, Dimitris G. Manolakis “*Digital signal processing (3rd ed.): principles, algorithms, and applications*,” Book. Prentice-Hall, Inc. Upper Saddle River, NJ, USA 1996 ISBN:0-13-373762-4
 - [46] R. Rondón, K. Landernäs and M. Gidlund, “*An analytical model of the effective delay performance for Bluetooth low energy*,” 2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Valencia, 2016.
 - [47] Z. Zhang, “*A flexible new technique for camera calibration*,” in IEEE Transactions on Pattern Analysis and Machine Intelligence, Nov 2000.
 - [48] S. Kokalj-Filipovic, L. Greenstein, B. Cheng and M. Gruteser, “*Methods for Extracting V2V Propagation Models from Imperfect RSSI Field Data*,” Vehicular Technology Conference (VTC Fall), 2015 IEEE 82nd, Boston, MA, 2015.
 - [49] A. Kita, P. Lorenzi, R. Rao and F. Irrera, “*Reliable and Robust Detection of Freezing of Gait Episodes With Wearable Electronic Devices*,” in IEEE Sensors Journal, 2017.
 - [50] G. Parthasarathy, A. Olmsted and P. Anderson, “*Natural language processing pipeline for temporal information extraction and classification from free text eligibility criteria*,” 2016 International Conference on Information Society (i-Society), Dublin, 2016.
 - [51] Online <https://www.astm.org/Standards/E1384.htm>
 - [52] Online <https://www.hl7.org/fhir/toc.html>
 - [53] Online: <http://fistarcatalogue.fiware.eng.it/enablers/ehr/documentation>
 - [54] M. S. Neofytou et al., “*Electronic Health Record Application Support Service Enablers*,” 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, 2015.
 - [55] F. De Cillis, F. De Simio, L. Faramondi, F. Inderst, F. Pascucci and R. Setola, “*Indoor positioning system using walking pattern classification*,” Control and Automation (MED), 2014 22nd Mediterranean Conference of, Palermo, 2014, pp. 511-516.
 - [56] A. Yaeli et al., “*Understanding customer behavior using indoor location analysis and visualization*,” in IBM Journal of Research and Development, vol. 58, no. 5/6, pp. 3:1-3:12, Sept.-Nov. 2014.
 - [57] G. Tang et al., “*Active tracking using color silhouettes for indoor surveillance*,” Wireless Communications & Signal Processing (WCSP), 2015 International Conference on, Nanjing, 2015, pp. 1-5.
 - [58] C. Xu, B. Firner, Y. Zhang, R. Howard and J. Li, “*Poster abstract: Exploiting human mobility trajectory information in indoor device-free passive tracking*,” Information Processing in Sensor Networks (IPSN), 2012 ACM/IEEE 11th International Conference on, Beijing, 2012, pp. 121-122.
 - [59] P. C. Liang, P. Krause, “*Real-time indoor patient movement pattern telemonitoring with one-meter precision*,” Wireless Mobile Communication and Healthcare (Mobihealth), 2014 EAI 4th International Conference on, Athens, 2014, pp. 141-144.

-
- [60] J. A. Kirkup, D. D. Rowlands, D. V. Thiel. “*Team Player Tracking Using Sensors and Signal Strength for Indoor Basketball*,” in IEEE Sensors Journal, vol. 16, no. 11, pp. 4622-4630, June1, 2016.
 - [61] H. Cho, W. Chung, “*Preliminary research on robust leg-tracking indoor mobile robots by combining the Kinect and the laser range finder information*,” Ubiquitous Robots and Ambient Intelligence (URAI), 2015 12th International Conference on, Goyang, 2015, pp. 462-464.
 - [62] M. Hedley, D. Humphrey, P. Ho, R. Shuttleworth, C. Mackintosh, “*Accurate tracking using TOA in sensor networks*,” Intelligent Sensors, Sensor Networks and Information Processing, 2008. ISSNIP 2008. International Conference on, Sydney, NSW, 2008, pp. 73-78.
 - [63] H. Xiong, Z. Chen, B. Yang, R. Ni, “*TDOA localization algorithm with compensation of clock offset for wireless sensor networks*,” in China Communications, vol. 12, no. 10, pp. 193-201, Oct. 2015.
 - [64] F. Wen, C. Liang, “*An Indoor AOA Estimation Algorithm for IEEE 802.11ac Wi-Fi Signal Using Single Access Point*,” in IEEE Communications Letters, vol. 18, no. 12, pp. 2197-2200, Dec. 2014.
 - [65] W. Ruan, Q. Z. Sheng, L. Yao, T. Gu, M. Ruta, L. Shangguan, “*Device-free indoor localization and tracking through Human-Object Interactions*,” 2016 IEEE 17th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM), Coimbra, 2016, pp. 1-9.
 - [66] S. Mahfouz, F. Mourad-Chehade, P. Honeine, J. Farah and H. Snoussi, “*Non-Parametric and Semi-Parametric RSSI/Distance Modeling for Target Tracking in Wireless Sensor Networks*,” in IEEE Sensors Journal, vol. 16, no. 7, pp. 2115-2126, April 1, 2016.
 - [67] M. H. Habaebi, R. O. Khamis, A. Zyoud, M. R. Islam, “*RSS Based Localization Techniques for ZigBee Wireless Sensor Network*,” Computer and Communication Engineering (ICCCE), 2014 International Conference on, Kuala Lumpur, 2014, pp. 72-75.
 - [68] Z. Li, D. B. Acuña, Z. Zhao, J. L. Carrera, T. Braun, “*Fine-grained indoor tracking by fusing inertial sensor and physical layer information in WLANs*,” 2016 IEEE International Conference on Communications (ICC), Kuala Lumpur, 2016, pp. 1-7.
 - [69] H. Yang; R. Zhan, J. Bordoy, F. Hoffinger, W. Li, C. Schindelhauer, L. Reindl, “*Smartphone based integrated Indoor Localization System using Inertial Sensor and Acoustic Transmitter/Receiver*,” in IEEE Sensors Journal , vol.PP, no.99, pp.1-1.
 - [70] S. Cai, W. Liao, C. Luo, M. Li, X. Huang, P. Li, “*CRIL: An Efficient Online Adaptive Indoor Localization System*,” in IEEE Transactions on Vehicular Technology , vol.PP, no.99, pp.1-1.
 - [71] H. Hashemi, “*The indoor radio propagation channel*” in Proceedings of the IEEE, vol. 81, no. 7, pp. 943-968, Jul 1993
 - [72] X. He, J. Li, D. Aloï, “*WiFi based indoor localization with adaptive motion model using smartphone motion sensors*,” 2014 International Conference on Connected Vehicles and Expo (ICCVE), Vienna, 2014, pp. 786-791.
 - [73] J. Yim, S. Jeong, J. Joo, C. Park, “*Utilizing Map Information for WLAN-Based Kalman Filter Indoor Tracking*,” Future Generation Communication and Networking Symposia, 2008. FGCNS '08. Second International Conference on, Sanya, 2008, pp. 58-62.

- [74] S. Feng, R. Murray-Smith, “*Fusing Kinect sensor and inertial sensors with multi-rate Kalman filter*,” Data Fusion & Target Tracking 2014: Algorithms and Applications (DF&TT 2014), IET Conference on, Liverpool, UK, 2014, pp. 1-8.
- [75] K. Lu, K. C. Chang, R. Zhou, “*The exact algorithm for multi-sensor asynchronous track-to-track fusion*,” Information Fusion (Fusion), 2015 18th International Conference on, Washington, DC, 2015, pp. 886-892.
- [76] Pengfei Guo, Xiaogang Wang, Yuehua Wu, “*Data Fusion Using Weighted Likelihood*,” European Journal of Pure and Applied Mathematics Vol. 5, No. 3, 2012, pp. 333-356.
- [77] E. Mirzazadeh, B. Moshiri, “Multi-sensor estimation using CKF and DKF,” Computer and Knowledge Engineering (ICCKE), 2013 3th International eConference on, Mashhad, 2013, pp. 321-325.
- [78] Sayed Amir Hoseini, Mohammad Reza Ashraf, “*Computational Complexity Comparison of Multi-sensor Single Target Data Fusion Methods by Matlab*,” International Journal of Chaos, Control, Modelling and Simulation (IJCCMS) Vol.2, No.2, June 2013.
- [79] Dirk Tenne, Tarunraj Singh. “*Optimal Design of Alpha-Beta-Gamma Filters*,” American Control Conference, 2000. Proceedings of the 2000 (Volume:6), pp 4348 – 4352.
- [80] C. Liu, A. Kiring, N. Salman, L. Mihaylova, I. Esnaola, “*A Kriging algorithm for location fingerprinting based on received signal strength*,” Sensor Data Fusion: Trends, Solutions, Applications (SDF), 2015, Bonn, 2015, pp. 1-6.
- [81] Gustavo Hernández-Peñaloza, Baltasar Beferull-Lozano, “*Field Estimation in Wireless Sensor Networks Using Distributed Kriging*,” IEEE ICC 2012 - Ad-hoc and Sensor Networking Symposium.
- [82] Shau-Shiun Jan, Shuo-Ju Yeh, Ya-Wen Liu, “*Received Signal Strength Database Interpolation by Kriging for a Wi-Fi Indoor Positioning System*,” Sensors 2015, 15(9), 21377-21393.
- [83] Tareq Alhmiedat, Ghassan Samara, Amer O. Abu Salem, “*An Indoor Fingerprinting Localization Approach for ZigBee Wireless Sensor Networks*,” European Journal of Scientific Research Vol. 105 No 2 July, 2013, pp.190-202.
- [84] Schucany, William R., “*Kernel Smoothers: An Overview of Curve Estimators for the First Graduate Course in Nonparametric Statistics*” Statist. Sci. 19 (2004), no. 4, 663–675.
- [85] F.-Á. Moreno, J. A. Merchán-Baeza, M. González-Sánchez, J. González-Jiménez, and A. I. Cuesta-Vargas, “*Experimental validation of depth cameras for the parameterization of functional balance of patients in clinical tests*,” Sensors (Basel), vol. 17, no. 2, p. 424, 2017.
- [86] Arsham Farshad, Jiwei Li, Mahesh K. Marina, Francisco J. Garcia, “*A Microscopic Look at WiFi Fingerprinting for Indoor Mobile Phone Localization in Diverse Environments*” 2013 International Conference on Indoor Positioning and Indoor Navigation, 28-31st October 2013.
- [87] Niu, J., Wang, B., Shu, L., Duong, T. Q., Chen, “*ZIL: An Energy-Efficient Indoor Localization System Using ZigBee Radio to Detect WiFi Fingerprints*,” IEEE Journal on Selected Areas in Communications, 33(7), pp. 1431-1442.
- [88] Beomju Shin, Jung Ho Lee, Taikjin Lee, Hyung Seok Kim, “*Enhanced Weighted K-Nearest Neighbour Algorithm for Indoor Wi-Fi Positioning Systems*,” 2012 8th International Conference on Volume 2, pp. 574 – 577.

- [89] Vahid Fathabadi, Mehdi Shahbazian, Karim Salahshour, Lotfollah Jargani. “*Comparison of Adaptive Kalman Filter Methods in State Estimation of a Nonlinear System Using Asynchronous Measurements*”. Proceedings of the World Congress on Engineering and Computer Science 2009.
- [90] Pengfei Guo, Xiaogang Wang, Yuehua Wu, “*Data Fusion Using Weighted Likelihood*,” European Journal of Pure and Applied Mathematics Vol. 5, No. 3, 2012, pp. 333-356.
- [91] G. Hernández-Peñaloza, C. Asensio-Marco and B. Beferull-Lozano, “*Distributed estimation of statistical correlation measures for spatial inference in WSNs*” Signal Processing Conference (EU-SIPCO), 2012 Proceedings of the 20th European, Bucharest, 2012, pp. 699-703.
- [92] Y. s. Chiou, C. I. Wang and S. C. Yeh, “*Reduced-complexity scheme using alpha-beta filtering for location tracking*” ET Communications, vol. 5, no. 13, pp. 1806-1813, Sept. 5 2011.
- [93] James Manyika, Michael Chui, Brad Brown, et al. “*Big data: The next frontier for innovation, competition, and productivity*,” <http://www.citeulike.org/group/18242/article/934132>
- [94] G. Tao, P. S. Archambault and M. F. Levin, “*Evaluation of Kinect skeletal tracking in a virtual reality rehabilitation system for upper limb hemiparesis*” 2013 International Conference on Virtual Rehabilitation (ICVR), Philadelphia, PA, 2013.
- [95] Wenjun Zeng. *Multimedia at Work* University of Missouri, Microsoft Research. 2012.
- [96] S. Sosa-Sesma and A. Perez-Navarro, “*Fusion system based on WiFi and ultrasounds for in-home positioning systems: The UTOPIA experiment*,” 2016 Int. Conference on Indoor Positioning and Indoor Navigation (IPIN).
- [97] X. Tian, R. Shen, D. Liu, Y. Wen and X. Wang, “*Performance Analysis of RSS Fingerprinting Based Indoor Localization*,” in IEEE Transactions on Mobile Computing, Oct. 1 2017.
- [98] M. Quan, E. Navarro, and B. Peuker, “*Wi-fi localization using rssi fingerprinting*,” California Polytechnic State University 1 Grand Avenue, San Luis Obispo, CA, Tech. Rep.
- [99] X. Tian et al., “*Improve Accuracy of Fingerprinting Localization with Temporal Correlation of the RSS*,” in IEEE Transactions on Mobile Computing, Jan. 1 2018.
- [100] Y. Xu, S. L. Lau, R. Kusber and K. David, “*An experimental investigation of indoor localization by unsupervised Wi-Fi signal clustering*,” 2012 Future Network and Mobile Summit (FutureNetw), Berlin, 2012.
- [101] Li Yan-Hua, Tang Wen-Sheng, Wang Sheng-Chun and Hui Peng, “*A New Method for Yielding a Database of Hybrid Location Fingerprints*,” Intelligent Computing Theories and Application: 12th International Conference, ICIC 2016, Lanzhou, China, August 2-5, 2016.
- [102] Kamol Kaemarungsi and Prashant Krishnamurthy, “*Properties of Indoor Received Signal Strength for WLAN Location Fingerprinting*,” Telecommunications Program, University of Pittsburgh.
- [103] Mohd Nizam Husen and Sukhan Lee, “*Indoor Human Localization with Orientation using WiFi Fingerprinting*,” Intelligent Systems Research Institute Sungkyunkwan University Republic of Korea.
- [104] C. Luo, H. Hong, M. C. Chan, J. Li, X. Zhang and Z. Ming, “*MPiLoc: Self-Calibrating Multi-Floor Indoor Localization Exploiting Participatory Sensing*,” in IEEE Transactions on Mobile Computing, Jan. 1 2018.

- [105] Qian Dong and W. Dargie, “*Evaluation of the reliability of RSSI for indoor localization*,” 2012 Int. Conference on Wireless Communications in Underground and Confined Areas, Clermont Ferrand, 2012.
- [106] R. Choudhary and H. K. Gianey, “*Comprehensive Review On Supervised Machine Learning Algorithms*,” 2017 International Conference on Machine Learning and Data Science (MLDS), Noida, 2017.
- [107] Soniya, S. Paul and L. Singh, “*A review on advances in deep learning*,” 2015 IEEE Workshop on Computational Intelligence: Theories, Applications and Future Directions (WCI), Kanpur, 2015.
- [108] L. Molina, T. Kerdoncuff, D. Shehadeh, N. Montavont and A. Blanc, “*WMSP: Bringing the Wisdom of the Crowd to WiFi Networks*,” in IEEE Transactions on Mobile Computing, Dec. 1 2017.
- [109] A. L’Heureux, K. Grolinger, H. F. Elyamany and M. A. M. Capretz, “*Machine Learning With Big Data: Challenges and Approaches*,” in IEEE Access, 2017.
- [110] M. Pak and S. Kim, “*A review of deep learning in image recognition*,” 2017 4th International Conference on Computer Applications and Information Processing Technology (CAIPT), Kuta Bali, 2017.
- [111] X. Wang, L. Gao and S. Mao, “*CSI Phase Fingerprinting for Indoor Localization With a Deep Learning Approach*,” in IEEE Internet of Things Journal, Dec. 2016.
- [112] X. Gan, B. Yu, L. Huang and Y. Li, “*Deep learning for weights training and indoor positioning using multi-sensor fingerprint*,” 2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Sapporo, 2017.
- [113] D. I. Nastac, F. A. Iftimie, O. Arsene, V. Ilian and B. Cramariuc, “*Indoor positioning WLAN based fingerprinting as supervised machine learning problem*,” 2017 IEEE 23rd International Symposium for Design and Technology in Electronic Packaging (SIITME), Constanta, 2017.
- [114] L. Yu, M. Laaraiedh, S. Avrillon and B. Uguen, “*Fingerprinting localization based on neural networks and ultra-wideband signals*,” 2011 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), Bilbao, 2011.
- [115] Antonio Estrada, Denis Efimov, Wilfrid Perruquetti, “*Position and velocity estimation through acceleration measurements*,” 19th IFAC World Congress, Aug 2014, Cape Town, South Africa. 2014.
- [116] L. Wang, Y. Sun, Q. Li and T. Liu, “*Estimation of Step Length and Gait Asymmetry Using Wearable Inertial Sensors*,” in IEEE Sensors Journal, May1, 1 2018.
- [117] Y. Kato, H. Nagano, M. Konyo and S. Tadokoro, “*Wearable gait logging system attached on ankles to estimate foot steps and trajectories*,” IEEE/SICE International Symposium on System Integration, Taipei, 2017.
- [118] Dan Murray, Ryan Bonick, “*Adaptiv: An Adaptive Jerk Pace Buffer Step Detection Algorithm*,” <https://github.com/danielmurray/adaptiv>.
- [119] Bluetooth technology, www.bluetooth.com/develop-with-bluetooth
- [120] IEEE 802.11 Wireless Local Area Networks, <http://www.ieee802.org/11/>

-
- [121] S. He, S. H. G. Chan, L. Yu and N. Liu, “*Maxlfd: Joint Maximum Likelihood Localization Fusing Fingerprints and Mutual Distances*,” in IEEE Transactions on Mobile Computing (Early Access)
 - [122] Munesh Singh, Pabitra Mohan Khilar, “*Actuating Sensor For Determining The Direction Of Arrival Using Maximal RSSI*,” International Journal of Scientific and Technology Research Volume 3, Issue 8, August 2014.
 - [123] M. Klemm and G. Troester, “*EM energy absorption in the human body tissues due to UWB antennas*”, Progress In Electromagnetics Research, 2006. doi:10.2528/PIER06040601
 - [124] H. Ahmadi and R. Bouallegue, “*Exploiting machine learning strategies and RSSI for localization in wireless sensor networks: A survey*”, 2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC), Valencia, 2017.
 - [125] Joaquín Torres-Sospedra, Raúl Montoliu, Adolfo Martinez-Usó et al. “*UJIIndoorLoc: A New Multi-building and Multi-floor Database for WLAN Fingerprint-based Indoor Localization Problems*,” Fifth International Conference on Indoor Positioning and Indoor Navigation, 2014.
 - [126] X. Wang, L. Gao, S. Mao and S. Pandey, “*DeepFi: Deep learning for indoor fingerprinting using channel state information*,” 2015 IEEE Wireless Communications and Networking Conference (WCNC), New Orleans, LA, 2015.
 - [127] B. J. Mohd, I. Amro and A. Alhasani, “*Indoor Wi-Fi tracking system using fingerprinting and Kalman filter*,” 2017 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), Aqaba, 2017.
 - [128] L. Khalil and P. Jung, “*Scaled Unscented Kalman Filter for RSSI-based Indoor Positioning and Tracking*,” 2015 9th International Conference on Next Generation Mobile Applications, Services and Technologies, Cambridge, 2015.
 - [129] N. Li, J. Chen, Y. Yuan and C. Song, “*A fast indoor tracking algorithm based on particle filter and improved fingerprinting*,” 2016 35th Chinese Control Conference (CCC), Chengdu, 2016.
 - [130] X. Tian, W. Li, Y. Yang, Z. Zhang and X. Wang, “*Optimization of Fingerprints Reporting Strategy for WLAN Indoor Localization*,” in IEEE Transactions on Mobile Computing, Feb. 1 2018.
 - [131] Weinberg, “*Using the ADXL202 in Pedometer and Personal Navigation Applications*,” Analog Devices, Inc.: Norwood, MA, USA, 2002.
 - [132] Tian Q., Salic Z., Wang K.I.-K., Pan Y., “*A Multi-Mode Dead Reckoning System for Pedestrian Tracking Using Smartphones*,” IEEE Sens. J. 2016.
 - [133] Bylemans I, Weyn M and Klepal M., “*Mobile phone-based displacement estimation for opportunistic localisation systems*,” In: 3rd international conference on mobile ubiquitous computing, systems, services and technologies (UBICOMM’09), Sliema, 11–16 October 2009, New York
 - [134] W. T. Higgins, “*A comparison of complementary and Kalman filtering*,” IEEE Trans. Aerosp. Electron. Syst., May 1975.
 - [135] D. Simon, “*Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*,” Hoboken, NJ, USA: Wiley, 2006.

- [136] R. Mahony, T. Hamel, and J.-M. Pfimlin, “Nonlinear complementary filters on the special orthogonal group,” IEEE Trans. Autom. Control, Jun. 2008.
- [137] S. O. Madgwick, “An efficient orientation filter for inertial and inertial/magnetic sensor arrays,” Dept. Mech. Eng., Univ. Bristol, Bristol, England, Tech. Rep. 25, 2010.
- [138] Y. Wu and W. Shi, “On Calibration of Three-Axis Magnetometer,” in IEEE Sensors Journal, Nov. 2015.
- [139] S. A. H. Tabatabaei, A. Gluhak and R. Tafazolli, “A Fast Calibration Method for Triaxial Magnetometers,” in IEEE Transactions on Instrumentation and Measurement, Nov. 2013.
- [140] Foram S. Panchal, Mahesh Panchal, “Review on Methods of Selecting Number of Hidden Nodes in Artificial Neural Network,” Int. Journal of Computer Science and Mobile Computing, Vol.3 Issue.11, 2014.
- [141] “S. B. Keser, A. Yazici and S. Günel, ”A hybrid fingerprint based indoor positioning with extreme learning machine,”, 2017 25th Signal Processing and Communications Applications Conference (SIU), Antalya, 2017
- [142] He, K., Zhang, X., Ren, S., Sun, J, “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification” In Proceedings of the IEEE international conference on computer vision 2015.
- [143] M. Sanjeev Arulampalam, Simon Maskell, Neil Gordon, and Tim Clapp, “A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking,” IEEE T. on Signal Processing, Feb. 2002.
- [144] Li Tiancheng, Bolic Miodrag, Djuric Petar, “Resampling Methods for Particle Filtering: Classification, implementation, and strategies,” Signal Processing Magazine (2015).
- [145] B. G. Sileshi, C. Ferrer and J. Oliver, “Particle filters and resampling techniques: Importance in computational complexity analysis,” 2013 Conference on Design and Architectures for Signal and Image Processing, Cagliari, 2013.
- [146] Martino Luca, Elvira Victor, Louzada Francisco, “Alternative effective sample size measures for importance sampling,” Conference: IEEE Workshop on Statistical Signal Processing (SSP) 2016.
- [147] K. Loumponias AND N. Vretos AND P. Daras AND G. Tsaklidis, “Using Tobit Kalman filtering in order to improve the Motion recorded by Microsoft Kinect,”, in Proceedings of the International workshop on applied probability IWAP 2016.
- [148] Madgwick, Sebastian OH and Harrison, Andrew JL and Vaidyanathan, Ravi. “Estimation of IMU and MARG orientation using a gradient descent algorithm,”, Rehabilitation Robotics (ICORR), 2011 IEEE International Conference on
- [149] W. S. Zheng and S. Gong and T. Xiang “Reidentification by Relative Distance Comparison”,, in IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013.
- [150] A. Møgelmoose and C. Bahnsen and T. B. Moeslund and A. Clapes and S. Escalera “Reidenti-
fication by Relative Distance Comparison”,, Tri-modal Person Re-identification with RGB, Depth and Thermal Features, 2013

- [151] F. Pala and R. Satta and G. Fumera and F. Roli. “*Multimodal Person Reidentification Using RGB-D Cameras*”, in IEEE Transactions on Circuits and Systems for Video Technology
- [152] A. Mignon and F. Jurie “*PCCA: A new approach for distance learning from sparse pairwise constraints*”, IEEE Conference on Computer Vision and Pattern Recognition, 2012.
- [153] M Shiomi, M Kurumizawa, T Kanda; H Ishiguro and N Hagita “*Finding a person with a Wi-Fi device in a crowd of pedestrians*”, Advanced Robotics, 2014
- [154] Masahiro Shiomi and Norihiro Hagita “*Finding a person with a wearable acceleration sensor using a 3D position tracking system in daily environments*”, Advanced Robotics, 2015.
- [155] K. Liu and Z. Zhao and A. Cai “*Parametric Local Multi-modal Metric Learning for Person Re-identification*”, , 2014 22nd International Conference on Pattern Recognition
- [156] K. Koide and J. Miura “*Person identification based on the matching of foot strike timings obtained by LRFs and a smartphone*”, 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
- [157] X. Tian, R. Shen, D. Liu, Y. Wen and X. Wang, “*Performance Analysis of RSS Fingerprinting Based Indoor Localization*”, in IEEE Transactions on Mobile Computing, vol. 16, no. 10, pp. 2847-2861, 1 Oct. 2017.
- [158] G. Papadopoulos and P. Daras “*Human Action Recognition Using 3D Reconstruction Data*”, , IEEE Transactions on Circuits and Systems for Video Technology
- [159] J. García and N. Martinel and G. L. Foresti and A. Gardel and C. Micheloni “*Person Orientation and Feature Distances Boost Re-identification*”, 2014 22nd International Conference on Pattern Recognition
- [160] Z. Gu, Z. Chen, Y. Zhang, Y. Zhu, M. L, and A. Chen “*Reducing fingerprint collection for indoor localization*”, Computer Communications, vol. 83, 09 2015.
- [161] K. Nedas *Implementation of Munkres-Kuhn (Hungarian) algorithm*, Algorithm, 2005
- [162] A. Gromek, R. Rytel-Andrianik, P. Samczynski and L. Corucci, “*Detection and tracking experiments with a fixed FMCW radar in an urban scenario*”, 2014 15th International Radar Symposium (IRS), Gdansk, 2014, pp. 1-5.
- [163] G. Gerstweiler, “*Guiding People in Complex Indoor Environments Using Augmented Reality*”, 2014 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Reutlingen, 2018, pp. 801-802
- [164] T. Pulkkinen, T. Roos, and P. Myllymäki, “*Semi-supervised learning for wlan positioning*”, in Proceedings of the 21th International Conference on Artificial Neural Networks - Volume Part I, ser. ICANN’11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 355–362.
- [165] S. Liu, H. Luo, and S. Zou, “*A low-cost and accurate indoor localization algorithm using label propagation based semi-supervised learning*”, in 2009 Fifth International Conference on Mobile Ad-hoc and Sensor Networks, Dec 2009, pp. 108–111.
- [166] S. He, S. H. G. Chan, L. Yu, and N. Liu, “*Fusing noisy fingerprints with distance bounds for indoor localization*”, in 2015 IEEE Conference on Computer Communications (INFOCOM), April 2015, pp. 2506–2514.

- [167] Y. Zhang, Y. Zhu, M. Lu, and A. Chen, “*Using compressive sensing to reduce fingerprint collection for indoor localization*,” in 2013 IEEE Wireless Communications and Networking Conference (WCNC)
- [168] J. Jun, L. He, Y. Gu, W. Jiang, G. Kushwaha, V. A, L. Cheng, C. Liu, and T. Zhu, “*Low-overhead wifi fingerprinting*,” IEEE Transactions on Mobile Computing, vol. PP, no. 99, pp. 1–1, 2017.
- [169] Y. Shu, Y. Huang, J. Zhang, P. Cou, P. Cheng, J. Chen, and K. G. Shin, “*Gradient-based fingerprinting for indoor localization and tracking*,” IEEE Transactions on Industrial Electronics, vol. 63, no. 4, pp. 2424–2433, April 2016.
- [170] J. Niu, B. Wang, L. Shu, T. Q. Duong, and Y. Chen, “*Zil: An energy- efficient indoor localization system using zigbee radio to detect wifi fingerprints*,” in IEEE Journal on Selected Areas in Communications, vol. 33, no. 7, pp. 1431–1442, July 2015.
- [171] S. Yoon, K. Lee, Y. Yun, and I. Rhee, “*Acmi: Fm-based indoor localization via autonomous fingerprinting*,” in IEEE Transactions on Mobile Computing, vol. 15, no. 6, pp. 1318–1332, June 2016.
- [172] A. Eleryan, M. Elsabagh, and M. Youssef, “*Synthetic generation of radio maps for device-free passive localization*,” in 2011 IEEE Global Telecommunications Conference - GLOBECOM 2011, Dec 2011, pp. 1–5.
- [173] D. Milioris, M. Bradonji, and P. Mhlethaler, “*Building complete training maps for indoor location estimation*,” in 2015 IEEE Conference on Computer Communications Workshops (INFOCOM WK-SHPS), April 2015, pp. 75–76.
- [174] M. Gunawan, B. Li, T. Gallagher, A. G. Dempster, and G. Retscher, “*A new method to generate and maintain a wifi fingerprinting database automatically by using rfid*,” in 2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Nov 2012, pp. 1–6.
- [175] F. Lemic, V. Handziski, G. Caso, L. D. Nardis, and A. Wolisz, “*Enriched training database for improving the wifi rssi-based indoor fingerprinting performance*,” in 2016 13th IEEE Annual Consumer Communications Networking Conference (CCNC), Jan 2016, pp. 875–881.
- [176] C. He, S. Guo, Y. Wu, and Y. Yang, “*A novel radio map construction method to reduce collection effort for indoor localization*,” Measurement, vol. 94, 08 2016.
- [177] Y. Cho, J. Kim, M. Ji, Y. Lee, and S. Park, “*Gpr based wi-fi radiomap construction from real/virtual indoor dynamic surveying data*,” in 2013 13th International Conference on Control, Automation and Systems (ICCAS 2013), Oct 2013, pp. 712–714.
- [178] Q. Li, H. Qu, Z. Liu, N. Zhou, W. Sun, J. Li, “*Af-dcgan: Amplitude feature deep convolutional gan for fingerprint construction in indoor localization system*,” 2018, [online] Available: <https://arxiv.org/pdf/1804.05347.pdf>.
- [179] Augustus Odena, Christopher Olah, Jonathon Shlens. “*Conditional Image Synthesis With Auxiliary Classifier GANs*,” [online] Available: <https://arxiv.org/pdf/1610.09585.pdf>
- [180] Bing Xu, Naiyan Wang, Tianqi Chen, Mu L., “*Empirical Evaluation of Rectified Activations in Convolutional Network*,” online Available: <https://arxiv.org/abs/1505.00853>

- [181] Nitish Srivastava and Geoffrey Hinton and Alex Krizhevsky and Ilya Sutskever and Ruslan Salakhutdinov. “ *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*”, Journal of Machine Learning Research, 2014- online Available: <http://jmlr.org/papers/v15/srivastava14a.html>
- [182] Haibing Wu and Xiaodong Gu “ *Max-Pooling Dropout for Regularization of Convolutional Neural Networks*”, online Available: <https://arxiv.org/pdf/1512.01400.pdf>
- [183] Mario Lucic, Karol Kurach, Marcin Michalski, Sylvain Gelly, Olivier Bousquet. “ *Are GANs Created Equal? A Large-Scale Study*”, Online Available: <https://arxiv.org/abs/1711.10337>
- [184] Xi Chen, Yan Duan, Rein Houthoofd, John Schulman, Ilya Sutskever, Pieter Abbeel. “ *InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets*”, Online Available: <https://arxiv.org/abs/1606.03657>

Figure 61: WGAN results on Fashion-MNIST Dataset for epochs 5, 25 and 50 respectively

Figure 62: GAF-WGAN results on Fashion-MNIST Dataset for epochs 5,25 and 50 respectively

Annex

9 Annex2

This annex provides the results for the GAF-GAN method applied to the rest of GAN architectures presented in chapter 6.

9.1 MNIST Dataset

9.1.1 Loss Function

9.2 CIFAR-10 Dataset

Figure 63: WGAN_GP results on Fashion-MNIST Dataset for epochs 5, 25 and 50 respectively

Figure 64: GAF-WGAN_GP results on Fashion-MNIST Dataset for epochs 5, 25 and 50 respectively

Figure 65: DRAGAN results on Fashion-MNIST Dataset for epochs 5, 25 and 50 respectively

Figure 66: GAF-DRAGAN results on Fashion-MNIST Dataset for epochs 5,25 and 50 respectively

Figure 67: Loss Function for ACGAN (*left*) and GAF-ACGAN (*right*)