Robust Expert Systems for more Flexible Real-World Activity Recognition

A dissertation submitted to

Universidad de Granada

for the degree of
Doctor of Science

presented by

Oresti Baños Legrán

MSc. Telecommunication Eng., University of Granada
born June 29, 1986
citizen of Spain

accepted on the recommendation of

Prof. Dr. Miguel Damas, UGR, co-director
Prof. Dr. Héctor Pomares, UGR, co-director
Prof. Dr. Ignacio Rojas, UGR, co-director

April 6, 2014

El doctorando D. Oresti Baños Legrán y los directores de la tesis D. Héctor Pomares Cintas, D. Ignacio Rojas Ruiz, Catedráticos de Universidad y D. Miguel Damas Hermoso, Profesor Titular de Universidad del Departamento de Arquitectura y Tecnología de Computadores de la Universidad de Granada

GARANTIZAMOS AL FIRMAR ESTA TESIS DOCTORAL

que el trabajo ha sido realizado por el doctorando bajo la dirección de los directores de la tesis y hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.

Granada, a 10 de marzo de 2014

Directores de la Tesis

D. Héctor Pomares Cintas D. Ignacio Rojas Ruiz D. Miguel Damas Hermoso

Doctorando

D. Oresti Baños Legrán

Acknowledgments

First and foremost, I wish to thank Prof. Ignacio Rojas, Prof. Héctor Pomares and Prof. Miguel Damas for their commitment and worth advising. They shared the support and enthusiasm in my work and provided optimal research conditions.

I would like to share the credit of this work with the students who contributed to this thesis through semester projects and Master Thesis as well as to thank Prof. Luis Javier Herrera, Prof. Juan Antonio Holgado, Prof. Alberto Guillén, Dr. Germán Sáez and Dr. Alberto Olivares as important research collaborators.

I wish to thank all the people I met during my research stays at ETH and TUe. Special thanks go to Prof. Daniel Roggen, Prof. Oliver Amft and Prof. Gerhard Troester for letting me be part of their outstanding groups and encourage me to be critic and analyst with my work. Likewise, I would like to specially thank Martin Kusserov and Alberto Calatroni for their insightful comments and help during the time I spent at ETH. Likewise, special thanks go to my TUe colleagues Gabriel, Florian, Amir and my officemates Iris, Aline and Máté.

Particular thanks go to my officemates and colleagues, formers and presents, Sandra, María, Nuria, Paloma, Curro, Fergu, Javi P., Javi A., Urquiza, Quique, Fran R., Fran B., Antonio M., Antares... whose most I consider more than fellows and brought an invaluable contribution to this thesis. I will not forget endless discussions and funny moments we all spent together.

I would like to thank my friends, Alvaro, Francis, Paco, Esteban, Diego and Juanma, and my friends from "El Laboratorio"', "Pante", "Bu", "Linda", "Maino", "Charles", "Perino", "Alfred", who helped me to destress and gave me unforgettable moments.

Last but not least, I most importantly wish to thank my parents Silvestre and Rafaela, my sister and husband Melania and Pepe, for their seamless support, advise and given love all long. Special heartfelt thanks go to Claudia, who made this thesis possible while wholeheartedly helping me. Without your support and love this would not have become true.

To those that I possibly (and unintentionally) leave out, please receive also my gratitude for any possible influence you could cause on me. To all them thanks for being a relevant boost to my career and life.

Table of Contents

Abstract

During the last years a tremendous interest in the analysis of human behavior has emerged to better understand and meet people's needs and demands. One of the most relevant research areas that investigate people behavior is human activity recognition. Human activity recognition aims at identifying human conducts in an autonomous fashion from the observation of a person's actions and their interaction with the surroundings. The flourishing of activity recognition is having a great impact in society and is extremely valuable in a wide variety of fields such as wellness, healthcare, sports or gaming.

The use of wearable or on-body sensors to monitor the human behavior is now on the forefront of human activity recognition. Nevertheless, the actual results for human activity recognition are fairly constrained and generally restricted to ideal or laboratory scenarios. Activity recognition systems are designed to comply with ideal conditions and are of limited utility in realistic domains. To become real-world applicable, activity recognition systems must satisfy operational and quality requirements that pose complex challenges, most of which have been sparsely and vaguely investigated to date.

Classic activity recognition systems assume that the sensor setup remains identical during the lifelong use of the system. However, in users' daily life, sensors may fail, run out of battery, be misplaced or experience topological variations. These changes may lead to significant variations in the sensor measurements with respect to the default case. Consequently, activity recognition systems devised for ideal conditions may react in an undesired manner to imperfect, unknown or anomalous sensor data. This potentially translates into a partial or total malfunctioning of the activity recognition system.

In this thesis, novel expert systems are proposed to address the challenges of making activity recognition systems functional in real-world scenarios.

An innovative methodology, the hierarchical weighted classifier, that leverages the potential of multi-sensor configurations, is defined to overcome the effects of sensor failures and faults. This approach proves to be as valid as other standard activity recognition models in ideal conditions while outperforming them in terms of robustness to sensor failure and fault-tolerance. This methodology also shows outstanding capabilities to assimilate sensor deployment anomalies motivated by the user

self-placement of the sensors.

Furthermore, a novel multimodal transfer learning method that operates at runtime, with low overhead and without user or system designer intervention is developed. This approach serves to automatically translate activity recognition capabilities from an existing system to an untrained system even for different sensor modalities. This is of key interest to support sensor replacements as part of equipment maintenance, sensor additions in system upgrades and to benefit from sensors that happen to be available in the user environment.

The potential of these advanced expert models leads to new research directions such as autonomous systems self-configuration, auto-adaptation and evolvability in activity recognition. Thus, this thesis opens-up a new range of opportunities for activity recognition systems to operate in real-world scenarios.

Resumen

El análisis del comportamiento humano ha suscitado un tremendo interés durante los últimos años. Este tipo de estudio se plantea como una herramienta fundamental para un mejor entendimiento de las necesidades individuales de cada persona así como para ayudar a satisfacer dichas necesidades. Uno de los campos de investigación con más relevancia dentro del estudio del comportamiento humano es el encargado del reconocimiento de la actividad humana. El reconocimiento de la actividad humana tiene como objetivo la identificación automática de las conductas del ser humano a partir de la observación de las acciones ejecutadas por el mismo y su interacción con el entorno que le rodea. El conocimiento adquirido a través del análisis del comportamiento humano es actualmente aprovechado en múltiples áreas, tales como el deporte o la industria del videojuego, manifestando un especial interés en el ámbito de la salud y el bienestar.

Los sistemas de monitorización portables o "vestibles" se encuentran a la vanguardia del reconocimiento de la actividad humana. Estos sistemas hacen uso de sensores capaces de medir el movimiento humano y que son integrados en dispositivos específicos o artículos de uso cotidiano como relojes, brazaletes o bandas. Son muchas las contribuciones proporcionadas hasta la fecha en el ámbito del reconocimiento automático de la actividad humana, alcanzando muchos de los sistemas propuestos una alta eficiencia en la detección e identificación de la actividad. No obstante, la inmensa mayoría de los sistemas desarrollados hasta la fecha son originariamente ideados para operar en condiciones ideales o de laboratorio. En consecuencia, el uso de estos sistemas en el mundo real presenta serias limitaciones que reducen de forma significativa su utilidad o incluso imposibilitan su aplicación. Para poder ser ampliamente utilizados en condiciones reales los sistemas de reconocimiento de la actividad requieren cumplir una serie de requisitos, los cuales presentan complejos retos que apenas han sido investigados en este campo.

Los sistemas de reconocimiento de la actividad son normalmente diseñados asumiendo una configuración predeterminada de los sensores. Más aún, se acepta que esta configuración permanece de forma invariable durante el uso habitual de dichos dispositivos. No obstante, los sensores están sujetos a fallos, errores o defectos consecuencia de caídas, roturas, falta de batería u otro tipo de anomalías tecnológicas.

Asimismo, el despliegue de los sensores puede cambiar sensiblemente durante el uso de los sistemas de reconocimiento. Este tipo de variaciones, perfectamente razonables en el mundo real, son difícilmente toleradas por aquellos sistemas concebidos para operar en condiciones ideales. En consecuencia, el uso de dichos sistemas en la vida real está sujeto a fallos que pueden llevar en algunos casos incluso a una total inoperancia del sistema.

En esta tesis se investiga el comportamiento de los sistemas de reconocimiento concebidos para operar en condiciones ideales cuando son utilizados en escenarios realísticos. Asimismo, esta tesis propone alternativas a la metodología de reconocimiento clásica para superar las limitaciones impuestas por el uso de dichos sistemas en el mundo real.

Un nuevo sistema experto, el clasificador jerárquico ponderado, capaz de aprovechar el potencial proporcionado por configuraciones multisensor se ha desarrollado expresamente para lidiar con los efectos producidos por fallos tecnológicos en los sensores. Dicho modelo demuestra capacidades de reconocimiento similares a las que proporcionan otros sistemas de reconocimiento estándar en condiciones ideales, además de superar a éstos en términos de tolerancia a fallos en los sensores. Asimismo, el clasificador jerárquico ponderado demuestra una alta tolerancia a variaciones introducidas en la disposición de los sensores en el cuerpo del sujeto, normalmente motivadas por la forma en que los usuarios se colocan los dispositivos que incorporan dichos sensores.

Otras variaciones en el sistema de sensado se pueden producir cuando sensores defectuosos u obsoletos son reemplazados por otros como parte de operaciones de mantenimiento, o cuando sensores adicionales a los definidos en fase de diseño son incorporados en un proceso de actualización del sistema de reconocimiento. En estos casos, resulta necesario realizar un entrenamiento específico para el uso de los nuevos sistemas incorporados, lo cual requiere de un proceso muy costoso atendiendo al modo de aprendizaje estándar para estos sistemas. Alternativamente, en esta tesis se define un nuevo método innovador que permite la transferencia automática de las habilidades de reconocimiento de la actividad de un sistema existente y funcional a otro que no cuenta con dichas capacidades o no está entrenado. El método desarrollado permite adems la transferencia de conocimiento multimodal, es decir, se puede utilizar incluso para la transferencia de conocimiento entre sistemas que operan sobre diferentes modalidades de sensores o fuentes heterogéneas de información. El proceso de transferencia de conocimiento se realiza de forma rápida y sin necesidad de intervención por parte del usuario

o un experto, lo cual lo hace especialmente viable para su utilización en contextos reales.

El potencial de los modelos expertos desarrollados en esta tesis abre por sí mismo nuevas líneas de investigación en el ámbito del reconocimiento de la actividad humana, como son los sistemas de reconocimiento auto-configurables y personalizables, los sistemas auto-adaptativos y los sistemas de reconocimiento evolutivos. En este sentido, esta tesis define un nuevo conjunto de oportunidades para optimizar el funcionamiento de los sistemas de reconocimiento de la actividad en el mundo real.

1

Introduction

1.1. Thesis goal

The goal of this thesis is to investigate on the effects of some of the most prominent technological and practical challenges posed by the use of on-body inertial sensing human activity recognition systems in the real-world. Moreover, this thesis aims at contributing with the development of robust expert models and systems especially devised to cope with these issues.

1.2. Activity recognition

Automatic recognition of human physical activities, in general activity recognition (AR), could be defined as the process of ascertaining the actions and goals of a person from a series of observations on the person's behavior and their surrounding conditions. Behavior observation and inference is normally performed through the use of computing and sensing systems that can be found on the person's area. Signal processing, machine learning and pattern recognition techniques are employed to capture the characteristics of people actions through the analysis of their body or corporeal movements. The identification of human actions is useful for a diverse range of applications that spans from healthcare or assistance to gaming, sports, human-computer interaction or industrial maintenance, among others.

Inference of human behavior could be approached through different sensing platforms depending on whether the sensors are placed on (wearable or on-body sensors) or around (ambient sensors) the person. The use of ambient sensors such as cameras or microphones is commonly restricted to scenarios on which their deployment is feasible. Consequently, AR capabilities are constrained to instrumented spaces and stationary settings, generally applying to indoor scenarios. Moreover, these technologies hardly meet important application and user requirements (e.g., privacy, occlusions, ambient noise) that make difficult their widespread use. On-body sensors lack of most of these limitations, which has contributed to put these at the forefront of the AR domain. On-body or wearable sensing promises ubiquitous and seamless recognition capabilities even for open-ended scenarios on which ambient sensing is impractical. From the sort of sensors that may be attached to the body, inertial sensors are the most exhaustively used, especially accelerometers. Accelerometers have been deeply characterized in other demanding domains such as aeronautics or automobile

industry, and can be now found in most mobile devices or embedded in daily-use accessories such as watches or shoes. Accelerometers are cheap, small and deliver rich information which make them specially suitable for activity-aware applications.

AR integrates sensor networks with machine learning techniques to model a wide range of human activities. The activity inference process consists of a set of stages on which signal processing and machine learning techniques are used. A set of sensors is deployed on the user's body to register their movements when performing a particular activity or encountering a specific situation. These sensors deliver raw unprocessed signals which represent the magnitude measured (e.g., acceleration). The registered information may be disturbed by electronic noise or other kind of artifacts. Depending on whether a certain information loss is tolerated, sometimes the signals are pre-processed through a filtering process. In order to capture the dynamics of the signals these are partitioned into segments of a fixed or variable size. Subsequently, a feature extraction process is carried out to enhance the characteristics unique to each activity and provide a more tractable representation of the signals for the pattern recognition stage. These features are provided as input to a classifier or reasoner, which ultimately yields the recognized activity to one of the considered for the particular devised problem. During the design phase, AR systems are trained on an annotated dataset of sensor readings on which experimental users are recorded while executing the set of target activities.

1.3. Challenges for real-world activity recognition

In the past few decades much research has been conducted on human activity recognition. Despite this, the take-up of the results has been fairly limited and generally restricted to ideal or laboratory scenarios. Laboratory systems are designed to comply with ideal conditions, however, real-world requirements are different. Consequently, AR systems developed under laboratory assumptions are of little utility in realistic domains. To become real-world applicable, AR systems must satisfy operational and quality requirements that pose complex challenges, most of which have been sparsely and vaguely investigated to date.

In the following, principal requirements and challenges of real-world AR systems are detailed.

Unobtrusiveness

Wearability limitations have traditionally been one of the most restrictive weaknesses in the real-world use of wearable computing. The origins of modern-day wearable computing could be attributed to the first prototypes developed at MIT [1], which consisted in a backpack-mounted computer to control cameras. Although pioneering everywhere personal computing, this kind of equipment was bulky, heavy, and not much ergonomic, thus far from being wearable but rather just portable. Astonishing technological advances have taken place since then, principally allowing for an impressive sensor miniaturization and cost reduction that are permitting to embed sensing technology and computing resources in a variety of items such as wrist-watches, glasses or other articles. Similarly, this technology may be also found in general purpose mobile devices, being smartphones the most broadly used exponent. This constitutes a huge step forward to make wearable systems convenient and acceptable for *people of the real-world*.

Although this new sort of systems are much more suitable than first wearable devices, not all people are used or willing to put on a wristwatch, wear bracelets or use glasses. No one can deny that systems that have not been part of our lives are now inconceivable out of them, as occur for smartphones, however, it is more likely to get users approval when technology gets part of their daily life in a soft manner. Specific of AR applications, systems must be comfortable and transparent to users to avoid constraining their actions as well as to not condition the way people behave in their daily lives. Smart-clothing is the paradigm devised to be the perfect means to support that. Nevertheless, smart-clothing poses complex challenges to be overcome. Apart from the difficulty of incorporating sensing technology to traditional apparel preparation procedures, major complexity is encountered to make this technology resistant and resilient to habitual processes such as washing, ironing and folding. In this regard, this technology must be weatherproof, waterproof, flexible and stainless, otherwise, be robust to real-world conditions.

Fashionability

There exist a stereotyped thinking of people carrying on wearable devices as technological “freaks” or “geeks”. This has been fairly motivated through the “extravagant” image given during the early stages of wearable computing. The use of huge sensing prototypes pictured a future closer to some film cyborgs and superheroes than systems

for real-world applications and people. Although some people may like this eccentric appearance, fashion plays an important role in our society, thus trendiness and stylishness must be characteristics to take into account when thinking of products to be commonly worn. Wearable technology companies are lately putting special attention to this fact, as demonstrates the incursion and alliances with several multinational corporations engaged in the design, development, manufacturing and worldwide marketing and selling of footwear, apparel and other type of accessories. An example of this new trend is the new generation of commercial items such as wrist-watches, bracelets or glasses that pursue to blend in with daily living wearables articles. Fashion may also help increase acceptability from those people that wish to preserve their privacy when wearing these systems. For example, AR systems could be utilized to help reduce childhood obesity, but users of dedicated systems could suffer from self-esteem or fear to be insulted or bullied in presence of other children. Therefore, a strong collaboration between technicians and fashioners is required to make these systems part of people lives, socially accepted and with which persons are familiar.

Usability

AR systems and applications must be defined to be controllable by ordinary people, which ranges from technological aware to non-expert users. Simple and understandable systems normally demonstrate to prevail over complex and cumbersome products. Therefore, sensor technologies and end-applications must be designed to minimize or even avoid the need of a specific users learning.

With respect to the sensing equipment, it must be a key aim of AR systems to guarantee as much flexibility as possible within the use of the sensor devices. Since the tendency is to encapsulate sensors in garments and accessories, it must be taken into account the way users normally utilize these elements in their daily living. To that end, experts must anticipate real-world situations that are normally disregarded during systems design and may have a tremendous negative impact in the correct functioning of AR systems. Typical in on-body motion sensing systems, these effects are principally observed when the sensor deployment varies with respect to their ideal or default distribution. Standard AR systems are not capable to cope with these variations, normally leading to a significant recognition worsening. For example, an AR system particularly designed to process the movements monitored on the right wrist (e.g., through a smartclock), will potentially fail to provide ac-

tivity detection when the sensor is worn on the left wrist. This kind of situation is quite recurrent in real-world settings, since as for this example some users may prefer to wear the wrist-watch on the left instead of the right wrist. Moreover, even when the sensor is placed in the correct limb or body part, nothing guarantees this will remain in that very position lifelong, not even during a short period of time. Loose-fitting garments or accessories are subject to placement drifts during their use (e.g., a bracelet moves from the wrist to the elbow). Moreover, users may wish to wear their clothes in a way different to the habitual fashion (e.g., rolled up sleeves). Standard AR systems are normally defined assuming that sensors are attached to the body in a firmly and tightly manner, thus potential errors could be expected under the event of sensor displacements. What must be clear is that these kind of assumptions are not applicable in the real-world, apart from going against important requirements such as unobtrusiveness and comfortability. From a design perspective these all are practical anomalies derived from the systems use and must be neatly addressed to guarantee an accurate and efficient activity assessment.

Privacy

Before the appearance of wearable systems, AR was mainly pursued through image and audio processing. In fact, numerous contributions have been provided during the last decades, with very promising results, but unfortunately there is a limitation within their use that seems to be impossible to overcome: privacy concerns. Recent experience demonstrates that people are not willing to let cameras or microphones get into their lives and observe what they normally do, independently of the benefit derived from their use. Conversely, on-body motion sensors surpass video and audio based systems since the recorded information is unfamiliar and looks noninvasive to the general user, thus not creating much controversy. In fact, people do not feel themselves to be observed but to be rather monitored, thus privacy concerns are almost nonexistent.

Yet, privacy concerns are more than reasonable since human behavior information could be maliciously utilized in case it ends up in the "wrong hands". It is true that many people actively share their life experiences through platforms such as social networks or blogs, possibly without knowing that some of this information could be used against them, but society is also becoming more and more technological aware and sensitive to the risks of sharing private information. Avoiding this

in the wearables domain could have important negative effects in the short-term, that may potentially lead to a generalized disbelief in the use of these systems. Therefore, it is important to address these issues in advance, ensuring mechanisms to fully protect users rights and privacy, as well as their safety.

Data-sharing

Another aspect that has not been deeply analyzed is the proprietary use of the data collected through monitoring devices. Commercial AR systems are normally devised for specific applications (e.g., fitness tracking, gaming). Therefore, if users wish to access diverse services, they need to wear multiple of these devices. For example, a user may wear a bracelet to monitor their workout, another bracelet to detect sleep disorders and a third one to determine their dietary habits, all devices from different vendors. As the number of applications or uses increases also the number of devices to be worn, which leads back to obsolete and obtrusive sensor configurations as the ones used in the origins of wearable computing. In many cases, the information captured through all these devices will be likely the same (e.g., motion of the lower arm). Therefore, the most reasonable option would be to have a set of general wearable devices capable of tracking users activities. The information collected through these systems could be leveraged by different applications as occurs with data collected through smartphones. In this regard, and to reduce to a bare minimum unobtrusiveness, smart-clothing is here seen again to be the most appropriate platform. This view opens-up a new market on which the focus is on the development of applications, while standard measurement platforms are shared for their use. Although a first attempt in this direction has being recently observed with projects such as Google Glass [2], there is much work to do to coordinate and generalize these ideas among application and hardware developers.

Simplicity

Most of the work in AR has been devoted to the task of achieving as much performance as possible in problems that account for an increasing number of subjects and activities. To that end, models of growing complexity are normally evaluated in lab settings and in an offline manner, thus utilizing computational resources that may not be available in real-world conditions. Although latest technical breakthroughs allow for higher processing and faster communications, these normally en-

tail important energy or battery demands, resources that are specially lacking in wearable computing. As the rest of sensing and processing units, batteries must also fit in shape and size with unobtrusiveness, lightness and wearability requirements, which seriously restricts their capacity. Wearable AR systems may be planned for an occasional usage, however, they are normally devised for an extensive daily utilization. Therefore, an optimal use of the supplies is required, specially for systems of long-term use. To reduce energy consumption AR systems should be simplified as much as possible. Nevertheless, to make systems simpler may entail a reduction of the recognition capabilities. Accordingly, a trade-off between simplicity and performance must be considered depending on the particular application requirements.

Latency

Activity detection and identification is not performed in an instantaneous fashion. Nevertheless, depending on the particular target application the inference of human behavior may be required at runtime or rather be more flexible in terms of time. Critical recognition systems such as the used to detect elderly falls should react as fast as possible to alert caregivers or trigger an emergency call to the nearest healthcare center. On the other hand, long-term monitors or trend analysis over long periods could be performed in a more relaxed basis (e.g., report on the calories burned during the last week).

As it happens to occur for simplicity and performance requirements, latency also interrelates with these. Simpler computational recognition models normally translates into faster activity detections, albeit this may reduce the accuracy of the recognizer. The use of distributed or parallel computing may also help reduce the reaction time for those cases in which the recognition process could not be simplified.

Fault-tolerance

Important advances have been performed during the last years to make sensor systems more efficient and perdurable. However, sensors are subject to degradation and damage due to their use, environmental changes or possible manufacturing defects. Under the event of degradation, the sensor data normally become corrupted or anomalous, therefore different to what is expected in normal conditions. At worst, sensors may break or stop working and no data explicitly delivered. Avoiding these technical issues has critical consequences on the normal behaving of standard recognition systems.

AR systems should be capable of detecting and dealing with sensor technical anomalies. Sensor anomalies detection is encountered of worth to inform and request specific maintenance. However, more importantly would be to make AR capable of handling these effects whether possible, specially for anomalies that are seen to appear occasionally or sporadically. For example, a sensor may provide anomalous data because it is not adequately supplied, however, this situation could be reversed when the battery is recharged. For such a task it makes no sense to request the help of a support team, and for sure not throw the sensor out. AR systems should be able to notice the importance of the sensor anomaly and use the rest of the sensor ecosystem to overcome the effects of this issue.

Critical sensor faults or breakdowns may normally require a complete replacement of the affected sensor. During maintenance tasks, AR systems should continue their work without interrupting the recognition process. AR models are classically defined to operate on predefined fixed sensor configurations, therefore they potentially need to be stopped until the equipment is restored. Some applications may not tolerate this type of interruptions, therefore recognition systems should implement mechanisms to continue their task albeit this may imply a worsening on the performance.

Self-configuration, auto-adaptation and evolvability

Configuration of AR systems is normally performed during the design phase, however, real-world recognition systems are subject to runtime changes hardly foreseeable during the systems development. Principal changes are associated to variations on the sensor setup (e.g., a new sensor introduced or an obsolete one replaced) or the activity concept (e.g., a new activity is included, the description of an existing one updated or unobserved activities removed from the original recognition set). Consequently, AR systems must be capable of adapting to real-world changing conditions of systems and users.

Like other systems, AR infrastructure is subject to upgrades and repairs. Moreover, different sensor configurations are envisioned during the course of a user's normal day. Depending on the particular context, users may wear specific garments (e.g., at work), casual clothes (e.g., at home) or specific accessories (e.g., at the gym). During these situations, sensors may be removed, substituted or newly added. Classic AR systems are trained on sensor data streams from datasets collected at design time with predefined and optimal sensor configurations.

Then, accounting for these variations would require to collect as many datasets as possible sensor configurations, which happens to be unfeasible. AR systems should implement mechanisms to autonomously adapt and self-configure their models to the actual sensing configuration. Likewise, these mechanisms should support the seamless integration of future sensing technologies with already existing ones, moving from constructive to evolvable paradigms. AR systems should be also defined to intelligently leverage those sensors that happen to be available to the user. Moreover, depending on the particular application, the use of part of the sensing infrastructure could be preferred. This may help to minimize energy and resources consumption, as well as reduce systems simplicity during execution time.

AR systems generally aim to work on a general basis, i.e., for diverse type of users. To that end, recognition models are built on data of various persons, seeking to learn the actual variability among subjects. Nevertheless, to capture the diversity among all people in the world is quite challenging, almost impossible, and certainly difficult to perform from the observation of a reduced set of experimental volunteers. To make this feasible, AR systems should learn the particular attributes of each user, thus adapting their parameters to their behavioral characteristics. To this end, a general model could be used as base system, and then evolve to fit with the user characteristics. Instead, systems could be directly personalized during the design phase, but developing a customized system for each person is unfeasible for obvious reasons. Even when personalization of systems could be procured, it does not ensure a lifelong operation of the system. Users characteristics and circumstances change during their life owing to diverse reasons (e.g., aging, varying health conditions), which further translates into a continuous drift on their normal behavior. In AR this means a change in the activities description with respect to the ones employed during the system design. This effect, also known as "concept drift", has to be taken into account to support the utilization of the AR system for the long-term, as well as to guarantee a certain level of recognition accuracy.

Systems need to dynamically and autonomously adapt to new situations that arise at runtime, evolving over time to operate in unforeseen circumstances during the initial design phase. However, for the adaptation process systems require extra knowledge or feedback, possibly gained from the analysis of their own behavior and the learning from past experiences. This feedback could come from external or internal sources. In recent works, external knowledge has been leveraged in

AR to improve systems efficiency and robustness, for example through asking users to notify whether the recognition system is making a detection error or not. Nevertheless, involving users in an active manner may generally result burdensome, which goes against systems usability requirements. Moreover, the information may be biased or erroneous because of unintentional misreports. The use of biofeedback (i.e., knowledge gained from the analysis of physiological functions) could be an appropriate means to avoid users explicit involvement, however, obtrusive interfaces are normally required (e.g., electroencephalogram caps for the detection of EEG error-related potentials). The best choice is possibly to make AR systems capable of providing its own internal feedback. Thus for example, recognition systems may exploit the knowledge gained from ensembles of decision makers, in which each sensor node provides a decision on the subject behavior and errors are detected from the divergence with respect to the majority opinion.

Reliability

Reliability is a decisive factor to determine the success of a product, specially for the long-term. Misleading users about AR system capabilities that are vaguely or even not supported could translate into a general disbelief in the use of these systems. In fact, day by day new AR commercial systems appear in the market claiming to provide "full" recognition capabilities, normally referring to the capacity of identifying any possible activity the user is able to perform. This is normally planned for marketing purposes, however, it does not reflect reality. Numerous challenges need to be still solved before making AR systems truly operable and reliable in the real-wold.

The accurate and precise detection of human activity poses several complex challenges. One of the principal challenges refers to the activity concept or definition. Although it could seem trivial at first sight, the complexity and diversity of human activities and actions make truly difficult to elaborate a clear definition of them. In fact, the way people interpret activities varies from person to person. Defining the start and end of an activity may be difficult (e.g., cooking starts when the user arrives to the kitchen or when the person holds the pan or when frying the steak?), activities may be interleaved (e.g., a user awaiting for the pan to get hot is just standing or also cooking?) or be of diverse granularity (e.g., standing still is an action, posture, gesture, activity?). Moreover, activities can be performed in many different ways, depending on who carries out the activity, in which conditions, environmental context,

etc. In human activity recognition using on-body inertial sensors, factors such as age, weight, height or other subject-related features, as well as ambient and context-related factors (e.g., subject carrying items, unstable floor, etc.) may determine that highly different data could refer to a similar activity concept. As an example, one cannot expect to register the same kind of data when an adult is cycling than when an elder does; similarly the gait may differ when walking on the ground, the grass or a frozen surface. Fortunately, there are characteristics on the habitual execution of some of these activities that in theory allow us to discriminate them. For example, depending on the intensity of the movements one should be able to distinguish between sedentary and dynamic activities. Nevertheless, some metrics may be useful for a group of people but not valid for others. For instance, speed could be used to differentiate between walking and running, however, the speed measured for an athlete when walking could possibly correspond to the top velocity of an elder while running.

Classic AR systems are trained on data collected in laboratory or ideal settings. People behavior in laboratory environments is normally conditioned because of their awareness and the influence of experts comments or presence. When applied to real-world conditions, a significant worsening on the systems operation is observed. To overcome this, real-world or naturalistic data should be used, thus helping overcome the gap between ideal and actual human behavior. In this regard, other important challenge in the AR field is to collect new datasets on which systems may be evaluated. Differently to other fields, there is no gold standard to universally validate the contributions. Therefore, there exists the need of a joint effort of the AR research community to collect rich general-purpose datasets as it occurs in other research fields (e.g, speech recognition, computer vision). With the lack of standardized datasets, the task of reproducing research turns to be quite difficult, which is found to be crucial in a research discipline.

Other challenges are faced during the learning of AR systems, such as data imbalancement or null class identification. Activity imbalancement refers to the fact that some activities are more frequently performed than others. For example, in long-term daily monitoring most of the time may be categorized in sleeping and working activities. Other activities may appear more occasionally or be completely infrequent. Models trained in an uneven amount of activity data could fail to generalize their recognition capabilities. Otherwise, some activities could be rigorously learned, while others may be weakly characterized. An-

other active area of investigation refers to the characterization of the null or "garbage" class. The null class comprises all those activities or actions that are not within the scope of the developed recognition system. Normally, only a few parts of the data stream are of interest to the AR systems (e.g., "data corresponding to gait"), thus there is a major part of the stream which is irrelevant (e.g., "all data not corresponding to gait"). This introduces once again an important imbalancement problem in which activities of interest (e.g., "walking") may be easily confused with activities of similar characteristics (e.g., "jogging or running"). These problems are shared with the more general field of pattern recognition, however, they must be also researched in the AR domain.

With regard to some claims of commercial AR products, we wish to specially notice that these systems are generally based on the use of a sole sensor unit. To ensure that "absolute" recognition capabilities could be achieved through the monitoring of a single body part is completely unrealistic. It is true that depending on the target application, the information registered through some body parts may be more valuable than from others. Nevertheless, common sense and experience allow us to state that the more diverse the target activities the more necessary a complete description of the body motion is required. To recognize all possible human activities (which is yet to be seen), information from the complete body is inevitably required.

Other important requirements such as an affordable price are not here specifically described but rather assumed. In fact, AR systems are envisioned to be applicable in many domains some of which are of limited resources, such as it occurs in the delivery of healthcare and remote assistance in developing countries. Therefore, and as a key factor of competitiveness, AR systems developers must keep in mind the use of realistic resources that do not excessively rise the price of the end-product.

1.4. Motivation and objectives

In the light of the challenges presented in Section 1.3 there is an opportunity to create more advanced systems capable of handling real-world AR issues as well as to incorporate more intelligent capabilities to transform experimental prototypes into actual usable applications. Thus the challenge facing this work is to identify and characterize some of the

most relevant limitations in the AR domain and develop solutions that may help overcome these complex problems.

The goal of this thesis is to investigate on the potential impact of some of the most prominent technological and practical issues in the use of on-body inertial sensing AR systems for the real-world, to demonstrate the limitations of classic solutions and to provide alternatives to cope with these effects. In this way, this work seeks to contribute to a better understanding of the needs of realistic activity-aware applications and aims to help paving the path to a new generation of AR systems readily available for their use in the real-world.

This thesis aims at achieving this goal via the following supporting objectives:

Objective 1: Investigate the tolerance of standard AR systems to unforeseen sensor failures and faults, as well as contribute with an alternate approach to cope with these technological anomalies.

Classic AR approaches assume that the sensor configuration remains identical during the lifelong use of the system. Nevertheless, as it happens to occur to any other electronic device, AR sensing systems are also subject to faults or technical anomalies normally due to harsh conditions. Changes in the sensor internal characteristics such as battery failures or sensor degradations may lead to potential variations in the sensor readings. Consequently, models trained on ideal signal patterns may react in a different fashion to imperfect or anomalous sensor data.

Major changes could be produced by critical sensor failures. At worst, sensors may get broken or damaged to an extent so they stop delivering data. Conceptually, similar situations may be observed when the user leaves a sensor behind or it is powered down, resulting in the loss of signals. Under these circumstances, AR models devised for ideal sensor configurations may potentially fail to provide activity-awareness capabilities.

Through this objective, the challenges posed by fault-tolerance requirements are aimed to be investigated. To that end, the robustness of standard AR systems to the effects of sensor functional anomalies should be evaluated and benchmarked with respect to their performance in ideal conditions. Likewise, this work also aims at contributing with an alternative AR model that may deal with these effects.

Objective 2: Research the robustness of standard AR systems to unforeseen variations in the sensor deployment, as well as contribute with an alternate approach to cope with these practical anomalies.

AR systems are normally designed to optimally operate on a specific sensor deployment. If the sensor distribution is modified with respect to the default deployment, the system recognition capabilities could severely worsen. In that case, sensor data streams and activity patterns registered from specific body parts and sensor orientations may potentially change when the sensors are distributed or worn differently to as originally planned. Therefore, standard AR systems learned on activity patterns of a default or ideal deployment may likely incur in misrecognition after changes in the sensor distribution.

Variations in the sensor deployment are principally motivated by the way users self-place or wear the sensors. Motion sensors are seen to be embedded into accessories or clothes, which may be worn on several different ways. Thus, a sensor devised to be worn on a specific body part could be accidentally or consciously placed in a different or even completely unrelated location (e.g., a smartclock is worn on the left instead of the right wrist, or a instrumented bracelet thought for the ankle is placed on the wrist). Likewise, sensors orientation may change when drifted or rotated (e.g., a wrist-watch pointing down instead of up). Placement and orientation changes could not only be originated when putting the sensors on but occur during the normal use of the systems. Thus, a sensor placed on a predefined target limb may drift during its use to a near position (e.g., a sport-belt devised for the chest may move to the upper abdominals during the exercising). Sensors embedded in loose-fitting garments are especially prone to experience these displacements in a continuous manner.

Sensor deployment changes not only apply to instrumented apparel or accessories but also to other articles. AR systems based on mainstream sensing technologies such as the implemented by mobile devices are also subject to deployment changes. For example, a smartphone could change its orientation from time to time, be originally placed in a shirt and then slip into a trousers pocket, or be held on a palm. Under all these real-world circumstances, embedded motion sensors will likely deliver different signals to what would be measured for a predefined or ideal sensor deployment.

Usability requirements make us necessary to take into account these potential situations and define more flexible models that may deal with

the effects of varying sensor deployments. Through this objective this thesis aims to analyze the effects of sensor changes derived from users self-placement, as well as investigate the impact of large depositionings that could be consequence of how people naturally wear and interact with instrumented items in realistic scenarios. In this regard, this work will also aspire to investigate the tolerance of classic AR systems to these variations and define models that may cope with these effects.

Objective 3: Study the capacity of standard AR systems to support unforeseen changes in the sensor network, as well as contribute with an alternate approach to cope with these topological variations.

Activity-aware systems are defined to be used for a specific sensor setup. Therefore, changing the sensors normally requires redesigning the system, which goes against durability and usability requirements. Sensor network variations refers to removal, replacement or inclusion of new sensors. Operationally, sensor removals are kind of similar to sensor breakages when these fail to deliver data, thus already considered in Objective 1. As a consequence, this objective rather focus on the challenge of substituting and adding sensors.

Sensor replacements are normally devised when a sensor breaks or reaches the end of its useful life. In this way, sensor substitutions are performed to recover the AR system default recognition capabilities. However, the new sensor may have different characteristics than the substituted one (e.g., different sampling rate, dynamic range), then the AR model might be likely incapable of leveraging the data measured through this. On the other hand, sensors could be newly incorporated to the default sensor topology (e.g., the user buys a new gadget or instrumented item). This poses a new sensor configuration that is normally unforeseen during the design phase. In both cases, a complete redefinition and retraining of the system is required, which may need to force the system to stop until the relearning is performed. Moreover, the learning process requires to collect new experimental data, which turns to be impractical for real-world settings.

Through this objective, some of the challenges posed by unobtrusiveness, usability and self-configuration requirements are investigated. Particularly, models to support sensor setup changes, i.e., inclusion of new devices or equipment replacement, avoiding specific maintenance and costly retraining of the systems are aimed to be developed.

1.5. Outline

This thesis is structured in six chapters.

Chapter 1 presents the goal of this thesis, introduces the paradigm of sensor-based human activity recognition, motivates research work in the field of real-world activity recognition, defines principal requirements and challenges related to this topic and details the supportive objectives to achieve the thesis goal.

Chapter 2 gives a concise overview of human activity recognition research, presenting the different stages of the activity recognition process and describing the principal methods and configurations utilized in on-body activity recognition. In a more detailed fashion, this Chapter also presents background in prior research and methods for dealing with sensor technological anomalies, sensor displacement and transfer learning in the activity recognition domain, which constitutes the three main areas of investigation of this thesis.

Chapter 3 investigates on the effects of sensor faults and failures on classic activity recognition approaches and presents and alternative method to cope with these technological issues.

Chapter 4 researches on the tolerance of activity recognition models to the effects of sensor displacement.

Chapter 5 investigates on the instruction of newcomer sensor systems and proposes an alternate method to facilitate the transfer of activity recognition capabilities between sensor systems.

Finally, in Chapter 6, achievements, contributions and final remarks of this thesis are presented. Ideas for possible future extensions of the presented research work are also drawn in this chapter.

2

State of the Art

2.1. Activity recognition chain

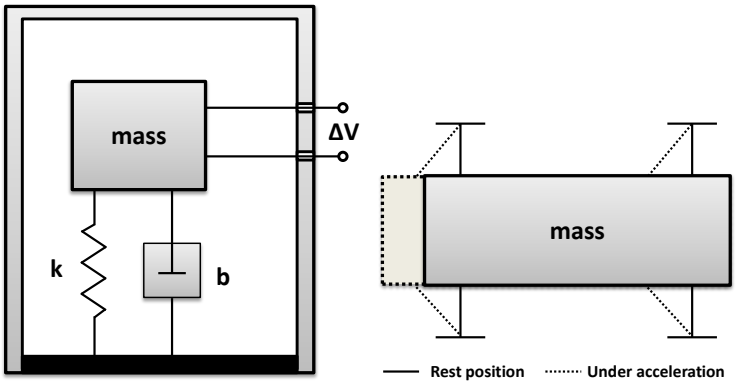
The activity recognition chain (ARC) [3], is the methodology most widely used in AR. The ARC consists of a set of specific steps that combines signal processing, pattern recognition and machine learning techniques to implement a specific AR system. Although most steps are normally implemented, some of these are of optional application. In the following the main stages of the ARC are described.

2.1.1. Signal acquisition

Firstly, the signals corresponding to the motion experienced by the on-body sensors are acquired. Diverse type of sensors may be used for this purpose, however, inertial measurement units (IMUs) such as accelerometers, gyroscopes and magnetometers are predominantly utilized. Accelerometers outstand among others due to their stability characteristics, robustness to environmental changes, reduced energy consumption and low cost. Moreover, these systems are highly miniaturizable and embeddable in diverse kind of devices and items. All this is possible due to a new generation of small devices called microelectromechanical systems (MEMS) which possess special properties that allow their integration in articles of the daily living.

Accelerometers produce voltage signals that are proportional to the experienced acceleration. There exist diverse techniques for transforming acceleration into an electrical signal. Generally, a mass is suspended on a linear spring from a frame which surrounds the mass (Figure 2.1(a)). When the frame is shaken as a consequence of the applied force, this begins to move further pulling the mass along with it. If the mass suffers the same acceleration as the frame, there needs to be a force exerted on the mass which leads to an elongation of the spring. Accelerometers of a miniaturized size, i.e., MEMS accelerometers, base on a slightly different structure. This consists of a suspended mass fixed to a substrate by suspension arms (Figure 2.1(b)). The experienced acceleration is proportional to the amount of displacement of the mass, which is measured through the tilt of the arms. Diverse displacement transducers could be used to measure this deflection, normally designating the accelerometer type. The most frequently used are capacitive, piezoelectric and piezoresistive accelerometers.

In order to measure accelerations in different directions, various accelerometers are simply positioned in an orthogonal manner. Biaxial or 2D-accelerometers (XY axes) and specially tri-axial or 3D-



(a) (b)

Figure 2.1: *General structure of moving mass accelerometers. (**a**) A mass is suspended from a frame on a linear spring with an elastic constant k and a resort with a damping constant b. The elongation of the spring is translated into a proportional voltage signal ΔV. (**b**) A mass is suspended from arms on a substrate. A displacement of the mass implies a tilt of the arms that are proportional to the experienced acceleration.*

accelerometers (XYZ axes) are the most utilized configurations for inertial sensing. Strictly speaking, a bi-axial accelerometer could be seen as the combination of two uni-axial accelerometers, as well as three uni-axial units define a tri-axial accelerometer. In fact, during the data collection a signal is particularly measured for each corresponding frame or axis. Therefore, in some works authors refer to this as a node or device that comprises various sensors. Nevertheless, in this dissertation the concept of sensor is rather used akin to node, i.e., the complete system capable of sensing, measuring, transducing and delivering the data associated to the motion experienced by a given body part.

In a nutshell, the acquisition process comprises the measurement of the physical phenomena (i.e., acceleration, inertia or movement suffered from the body part on which the sensor is mounted), conversion to electrical signals (i.e., transduction of the movements of the seismic mass) and encoding into machine-readable digital data that can be processed through computers (i.e., analog to digital converter). A complete description of the operation process may be seen in [4].

2.1.2. Signal preprocessing

Secondly, the registered data is preprocessed to remove noise and diverse type of artifacts such as spurious spikes or electronic noise, typically through a filtering process. Commonly used filters for acceleration signals are median, Butterworth low/high-pass, discrete wavelet, Wiener filters and Kalman filters.

Preprocessing of acceleration does not only comprise signal filtering but involves, among others, calibration of the sensors (e.g., procure axes orthogonality, correct misaligment between body and sensor frames), unit conversion (e.g., from 'mV' to 'G' or 'm/s^2'), synchronization or resampling/downsampling. Most of this processing is directly performed on a hardware basis, however, other techniques are applied through specific or general computation. A review on preprocessing techniques for AR from accelerometer data is presented in [5, 6].

Some applications also use preprocessing techniques to separate the different components of the signal, for example into dynamic and static acceleration components. Dynamic or linear acceleration component represents the acceleration due to body motion while the static or gravitational acceleration component relates to the force exerted by the Earth's gravity.

2.1.3. Signal segmentation

The preprocessed data stream is subsequently partitioned into segments of a certain length to capture the dynamics of the signals. This process has been performed in different ways in the AR field. Most of the segmentation techniques could be categorized in three groups, namely activity-defined windows, event-defined windows and sliding windows. The activity-defined windowing procedure consists in the partitioning of the sensor data stream based on the detection of activity changes. Initial and end points are determined for each activity, prior to explicitly identifying the specific activity. The event-defined approach bases on locating specific events which are further used to define successive data partitioning. Since the events may not be uniformly distributed in time the size of the corresponding windows is not fixed. In AR normally identified events are heel strikes or toe-offs typically used in gait analysis. Finally, in the sliding window approach, the signals are split into windows of a fixed size and with no inter-window gaps. A wide range of window sizes have been used in previous studies from 0.1s [7] to 12.8s [8] and even more [9]. An overlap between adjacent windows is toler-

ated for certain applications [10, 11], however, this is less frequently used. The sliding window approach is the most widely employed segmentation technique in AR. Its implementational simplicity and lack of preprocessing determines it as ideally suited to real-time applications.

2.1.4. Feature extraction and dimension reduction

Next, a characterization of each data segment is performed. The feature extraction process is carried out to provide a more tractable representation of the signals for the pattern recognition stage. A wide range of heuristics [12, 13], time/frequency domain [14, 15, 16], wavelets coefficient [17, 18] and other sophisticated mathematical and statistic functions are used (see review in [18]).

In some cases, a feature selector is used to reduce redundancy among features as well as to minimize the feature space dimension. In fact, the higher the dimension of the feature space is, the more computationally intensive the reasoning process turns to be. Minimizing computational power, memory and bandwidth requirements is specially sought for embedded systems for real-time AR. An important objective when selecting features is to try to maintain the desired target performance, however, sometimes a trade-off must be found. Examples of feature selection methods used in AR are principal or independent component analysis [19], forward-backward selection [7] or correlation [15]. The literature provides many more techniques that are used in other fields, such as branch and bound, best first, beam search or relief. More information about feature selection techniques could be found in [20].

2.1.5. Classification

Finally, the resulting feature vector is provided as input to a reasoning or classification model, which eventually yields the recognized activity. Classifiers identify to which of a set of classes a new observation or instance belongs, on the basis of a training set of data containing instances whose category membership is known. A wide variety of machine learning, pattern recognition and data mining techniques is normally used for this purpose. This includes from models devised for capturing time dependencies and sequences (hidden Markov models, HMM) [21, 22], fuzzy models operating on rules and membership functions for approximation [23], to techniques leveraging the potential of ensembles of weak classifiers such as boosting and its variants [24, 25].

Nevertheless, the most broadly used are classical or standard supervised techniques. Examples of these techniques are decision trees (DT) algorithms which examine the discriminatory ability of the features one at a time, creating a set of rules that ultimately leads to a complete classification system. DT proved to perform well in combination with time and frequency domain features [10, 15, 26], although they showed less accurate for other setups [27]. Due to its simplicity, speed and absence of a training phase, the k-nearest neighbor (KNN) is also one of the most used techniques in machine learning. Based on a neighborhood majority voting scheme, the classification of a given sample is assigned to the most common class amongst its K nearest neighbors. Interesting results have been showed from its use in [7, 18, 28]. Another frequently used approach is the based on the Bayes' rule. The naive Bayes (NB) algorithm can be a suitable approach as long as the stochastic independence is guaranteed, which in practice is normally attained. This technique has been successfully used in prior AR problems [16, 29, 15]. Support vector machines (SVM) is another standard learning technique which has become very popular in the last years. The promising results recently obtained in previous studies as [30] or [31] reinforce the idea of its use.

2.1.6. Operation of the ARC

The ARC could be applied in two different modes of operation, namely training or modeling and classification. Given a set of categorical labels (ground truth) associated to a set of entries (inputs), the classification task consists in assigning these entries (prediction) into one of the predefined classes. In training mode, the features extracted in a previous stage and the corresponding ground truth class labels are used as input to train a classifier model. In classification mode, the features and the previously trained model are used to calculate a score for each activity class and to map these scores into a single class label in the classification stage.

The classification may be of a binary (two activities) or N-ary (N activities) nature. In the N-ary or multi-class scenario, the input is normally to be classified into one, and only one, of N non-overlapping classes. In the binary case (N=2), the input is to be classified into one, and only one, of two non-overlapping classes.

2.1.7. Performance and metrics evaluation

The evaluation of AR systems is normally carried out through a cross-validation process. In AR two cross-validation methods are predominantly utilized: leave-one-subject-out cross-validation (LOOXV) and ten-fold cross-validation (10-fold XV). As summarized in [32] and according to [33, 34], LOOXV is the best technique for risk estimation whereas 10-fold XV is the most accurate approach for model selection, which applies well for models comparison. Accordingly, and since this work intends to compare the capabilities of multiple AR models, in this thesis the evaluation is performed through the 10-fold XV model. Moreover, the cross-validation process should be normally repeated a sufficient number of times to ensure statistical robustness as well as to procure an asymptotic converge to a correct estimation of the systems performance [35].

To evaluate the performance of AR systems diverse metrics may be used. The confusion matrix stands out among others since it collects all the information corresponding to the performance evaluation in a single matrix (Figures 2.2-2.3). From this matrix, several other metrics could be simply derived.

For the binary classification (Figure 2.2) the system correctness is assessed by computing the number of correctly recognized class examples (true positives, TP), those that do not belong to the class (true negatives, TN), the number of examples incorrectly assigned to the class (false positives, FP) and the ones not recognized as class examples (false negatives, FN). In the N-ary case (Figure 2.3) only the diagonal elements corresponds to correct classifications (tp_k) while the off-diagonal comprises all missclassifications (ϵ_{ij}).

		Inferred Classes		
		Positive	Negative	
Actual Classes	Positive	TP	FN	$A_{Pos} = TP + FN$
	Negative	FP	TN	$A_{Neg} = FP + TN$
		$I_{Pos} = TP + FP$	$I_{Neg} = FN + TN$	$Q_{Bin} = I_{Pos} + I_{Neg} = A_{Pos} + A_{Neg}$

Figure 2.2: *Confusion matrix for a binary problem (a.k.a, diagnostic or contingency table). The matrix contains the true positives (TP), false positives (FP), false negatives (FN) and true negatives (TN) scores.*

		Inferred Classes				
		1	2	$\cdots$	N	
Actual Classes	1	tp_1	ϵ_{12}	$\epsilon_{1\cdots}$	ϵ_{1N}	$A_1 = tp_1 + \sum_{j\neq 1}^{N} \epsilon_{1j}$
	2	ϵ_{21}	tp_2	$\epsilon_{2\cdots}$	ϵ_{2N}	$A_2 = tp_2 + \sum_{j\neq 2}^{N} \epsilon_{2j}$
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	N	ϵ_{N1}	ϵ_{N2}	$\epsilon_{N\cdots}$	tp_N	$A_N = tp_N + \sum_{j\neq N}^{N} \epsilon_{Nj}$
		$I_1 = tp_1 + \sum_{i\neq 1}^{N} \epsilon_{i1}$	$I_2 = tp_2 + \sum_{i\neq 2}^{N} \epsilon_{i2}$	$\cdots$	$I_N = tp_N + \sum_{i\neq N}^{N} \epsilon_{iN}$	$Q = \sum_{j=1}^{N} A_j = \sum_{i=1}^{N} I_i$

Figure 2.3: *Confusion matrix for a N-ary problem. Classification errors between classes correspond to the off-diagonal elements (ϵ_{ij}) while correct classifications are given in the matrix diagonal (tp_k).*

From all available metrics, accuracy is the most widely used. The accuracy rate is defined as the proportion of correct classifications ($TP + TN$ in the binary case, $\sum tp_k$ in the N-ary case) with respect to the total classified instances (Q_{bin} in the binary case, Q in the N-ary case). The accuracy is a simple metric, although may suffer from some limitations under the presence of data imbalancement [31], which may be specially relevant in problems with a reduced class set. Sensitivity (Se) and specificity (Sp) are metrics normally used in those cases. Sensitivity defines the proportion of positive results that are correctly recognized ($Se = TP/(TP + FN)$), while specificity measures the ability to identify negative results ($Sp = TN/(TN + FP)$). Other metrics such as F-measure, G-mean, AUC, ROC or confusion entropy are proposed in other fields [36, 31], but seldom used in AR evaluations.

2.2. Standard AR models

Depending on the employed sensor topology, diverse variants of the general ARC may be devised. In this Section, the three most predominant variants of the ARC are presented. The first one corresponds to the case in which a single sensor unit is utilized. The second and third models apply to multiple sensor setups.

2.2.1. Single sensor ARC

The simplest ARC configuration is defined for the case in which a sole sensor is used to register body motion. Hereafter called single sensor activity recognition chain (SARC), the description provided in Section 2.1 directly applies to this configuration. In Figure 2.4 a scheme of the complete SARC is depicted. In short, a sensor S delivers a stream of raw signals (u) corresponding to the motion measured for a specific body part (e.g., 'X-axis and Y-axis acceleration measured on the wrist'). These signals are preprocessed before segmentation. The preprocessed data (p) are subsequently k-partitioned (s_k) and a set of features (generically defined as f) are extracted from these. The feature vector ($f(s_k)$) is used as input to a trained classifier, which eventually yields the recognized activity (c).

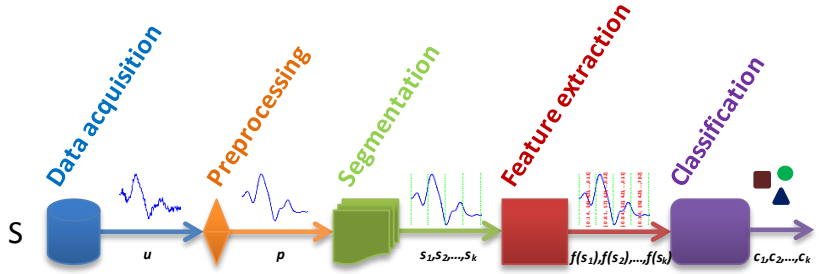


Figure 2.4: *Single sensor activity recognition chain (SARC). A sensor S delivers raw signals (u) which may be optionally preprocessed (p). The signals are k-partitioned (s_k) and a set of features (generically defined as f) are extracted from these. The feature vector ($f(s_k)$) is used as input to the classifier entity, which yields an activity or class c on a N-class problem.*

Many studies have explored the recognition of human activities through the use of a single sensor [13, 16, 12, 37, 38, 39, 8].

2.2.2. Multi-sensor ARC

The monitoring of various body parts require the use of multiple sensor devices. When several sensors are utilized the normal approach is to partially or completely replicate the diverse stages of the ARC for each sensor node. Therefore, given a body sensor network (BSN) com-

posed by M sensors, the system comprises M individual ARCs, i.e., a multiple sensor activity recognition chain (MARC). Specific to this approach, acquisition, preprocessing, segmentation and feature extraction stages are individually executed for each sensor stream. As a result, a feature vector is provided for each individual sensor. At this point, two approaches may be followed, either these M independent feature vectors are combined in a way to be employed as input to a single classification entity (see Section 2.2.3) or respectively used as inputs to a set of M standard classifiers (see Section 2.2.4).

2.2.3. Feature fusion multi-sensor ARC

Feature fusion, in advance defined as feature fusion multi-sensor activity recognition chain (FFMARC), consists in the combination or aggregation of the feature vectors computed for each individual sensor [40]. The resulting aggregated feature vector is used as input to a single classifier. Strictly speaking, fusion of multiple signals is normally performed in each individual ARC. Features extracted from each sensor frame (e.g., X axis and Y axis) are normally combined in a unique feature vector that characterizes the motion measurements coming from that node. Nevertheless, this is actually rather seen as a feature (e.g., 'standard deviation of the acceleration measured in the X axis'). Therefore, the concept of fusion is here applied to the aggregation of the information collected from various body parts through diverse sensor devices.

The FFMARC operates as shown in Figure 2.5. M sensors deliver raw signals $(u_1, u_2, ..., u_M)$ which are subsequently preprocessed $(p_1, p_2, ..., p_M)$. The preprocessed signals are partitioned into data windows of a given length $(s_{1k}, s_{2k}, ..., s_{Mk})$ and a set of features are extracted from these, possibly different for each chain $(f_1, f_2, ..., f_M)$. For a specific time window k, the M feature vectors obtained across all sensor streams are combined into a single feature vector $(\{f_1(s_{1k}), f_2(s_{2k}), ..., f_M(s_{Mk})\})$. The resulting feature vector is used as input to a trained classifier, which eventually yields the recognized activity (c_k).

FFMARC is possibly the most widely used approach in AR for multi-sensor configurations. It has been for example utilized in [19, 41, 42, 43, 10, 15].

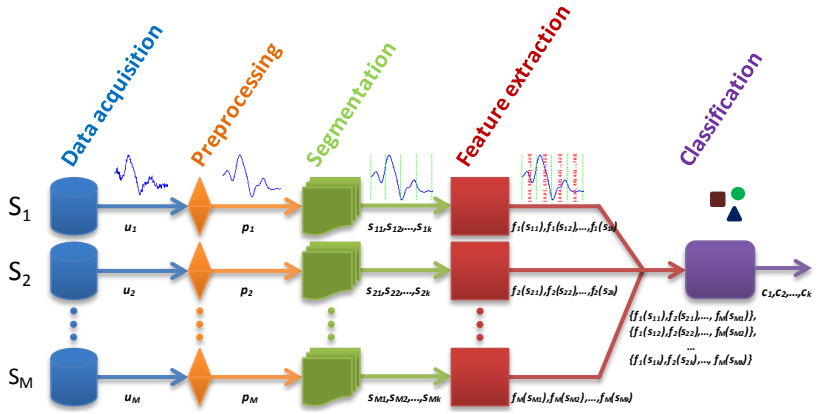


Figure 2.5: *Feature fusion multi-sensor activity recognition chain (FFMARC). M sensors deliver raw signals $(u_1, u_2, ..., u_M)$ which are subsequently processed $(p_1, p_2, ..., p_M)$. The signals are k-partitioned $(s_{1k}, s_{2k}, ..., s_{Mk})$ and a set of features extracted from each data window. For each window k, the feature vector computed across all sensors is aggregated into a single feature vector $(\{f_1(s_{1k}), f_2(s_{2k}), ..., f_M(s_{Mk})\})$ that is used as input to a classifier. The classifier yields an activity or class c on a N-class problem.*

2.2.4. Decision fusion multi-sensor ARC

AR for multi-sensor configurations may be also performed through the combination of the decisions delivered by independent ARCs each one associated to a sensor. Decision fusion, here defined as decision fusion multi-sensor activity recognition chain (DFMARC), is shown in Figure 2.6. M sensors deliver raw signals $(u_1, u_2, ..., u_M)$ which are subsequently preprocessed $(p_1, p_2, ..., p_M)$. The signals are segmented in data windows of a given length $(s_{1k}, s_{2k}, ..., s_{Mk})$ and a set of features are separately extracted from each data window. For a data window k, the extracted feature vectors are used as inputs to the corresponding classification entity (e.g., $f_1(s_{1k})$ for $S1$, $f_2(s_{2k})$ for $S2$, etc.). Every individual classifier yields a class c on a N-class problem; thus M decisions are in overall provided, one per sensor $(c_1, c_2, ..., c_M)$. To provide a unique recognized class, each sensor decisions are further combined through a fusion model (ψ), which eventually yields the recognized activity or class $(c = \psi(c_1, c_2, ..., c_M))$.

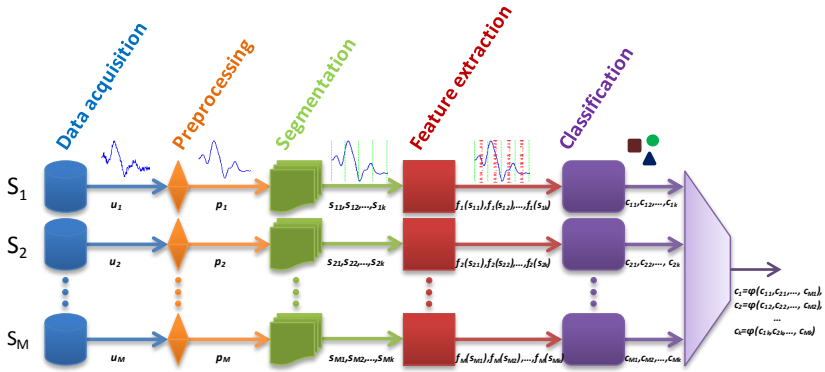


Figure 2.6: *Decision fusion multi-sensor activity recognition chain (DFMARC). M sensors deliver raw signals $(u_1, u_2, ..., u_M)$ which are next preprocessed $(p_1, p_2, ..., p_M)$. The signals are k-partitioned $(s_{1k}, s_{2k}, ..., s_{Mk})$ and a set of features extracted from each data window k $(f_1(s_{1k}), f_2(s_{2k}), ..., f_M(s_{Mk}))$. Each feature vector is used as input to the classifier entities. Each classifier yields a class c on a N-class problem, which are further combined through a decision fusion technique to provide the eventual recognized class $(c = \psi(c_1, c_2, ..., c_M))$.*

There exist multiple aggregation models [44, 45], however, this work particularly considers some of the most widely used. Concretely, hierarchical decision (HD) and majority voting (MV) schemes are in the following presented.

Hierarchical decision (HD)

Let us consider a set of classifiers each one operating on the data extracted from a specific sensor. Depending on the target activity, the information captured through some of these sensors may be more valuable than through others. Therefore also the decisions provided by each classifier may have different relevance. Then, the idea is to give more importance to those classifiers that generally behave better, thus allowing them to decide first. From this, decisions are made in strict order of classification capabilities (the ranking is normally established according to performance criteria). Thus, if a top ranked classifier positively recognizes the activity, this would be provided as the eventual decision. On the contrary, if the class is rejected, the next classifier in the

hierarchy is asked and so on.

One of the main drawbacks of the HD model is the dependence of the low-level entities decisions on the upper ones. If a better ranked classifier fails, the error propagates down through the hierarchy, potentially leading to a misclassification. It is true that hierarchical configurations in principle minimize the likelihood error when there exist large reliability differences between top and weak classifiers. However, this could turn against when a slight gap exists among the performances of each decision entity. In that case, leaving reliable classifiers out of the decision making is encountered to be a loss. HD is a model that behaves reasonably well when its top entities are reliable enough, but may miss the classification potential of the middle- and low-ranked decision entities.

Hierarchical decision models have been utilized in [46, 47, 48, 49, 26, 27, 50] for AR purposes.

Majority voting (MV)

Majority voting or plurality voting is a simple approach that relies on an equality scheme. This model, based on a democracy-approach, give the same opportunities to all decision entities. The eventual adopted decision is the one obtaining more votes from the participant decision entities. The main properties of this method are fairness and decisiveness, which translate into a similar treatment of each vote. These properties may be specially sought when rich sensor topologies are considered.

Nevertheless, MV advantages may also constitute its main limiting drawbacks. Similar importance is given to all decisions even when they may be differently accurate. Consequently, a relevant performance worsening could be expected when a majority of the decision makers are of weak or poor nature ("tyranny of the majority").

Majority voting models have been utilized in [51, 52, 53, 54, 55, 56] for AR purposes.

2.3. Sensor faults and failures in AR

Sensor devices are subject to technological faults and failures that may affect the normal functioning of the AR systems. While sensor failures lead to an interruption of the sensor data streaming, faulty sensors do not completely fail to continue reporting values, although these normally have a new meaning or are potentially invalid. A complete

taxonomy and review of common sensor faults and failures in sensor networks is presented in [57].

Several techniques have been developed for the detection of node failures in sensor networks. One of the most utilized approaches consists in continuously querying the network nodes to identify sensors that fail to deliver data [58, 59]. These methods are quite practical for sensor failure detection but of little utility when detecting faulty sensors. Accordingly, other more sophisticated techniques have been provided in the literature to detect faulty sensors. Ramanathan *et al.* [60] exploit the correlations between the data of neighboring sensors. In [53] the authors utilize the correlation changes encountered between data streams corresponding to the observation of a related physical phenomena. Also at the signal level, Yao *et al.* [61] evaluate the similarity between data streams to simply and efficiently identify sensor faults. These techniques do not necessarily require to know the specific meaning of the sensor data and may operate on the raw sensor signals. A comparison among the measures of neighbor sensors is also carried out in [62]. Here a probabilistic approach is particularly proposed to determine the status of a sensor depending on the decisions of the surrounding sensors. In [60] the authors detect sensor faults through the identification of outliers at the feature level. Fault diagnosis has also been performed through the use of clustering [63] and deterministic learning [64] techniques utilized in a distributed manner to reduce communication overhead and minimize energy consumption. Ganeriwal *et al.* [65] utilize reputation systems to develop a community of trustworthy sensor nodes at runtime based upon the behavior of these nodes. They provide a scalable, diverse and generalized approach to counter all types of misbehavior resulting from faulty nodes in a sensor network.

To overcome the effects of sensor faults and failures diverse approaches have been proposed. For general sensor networks, the use of back-up sensor systems has been for example explored in [66]. Data imputation techniques have been proposed in [67] to substitute sensor missing values through artificially generated data. Similarly, regression models have been recently utilized in [68] to predict and replace loss sensor data. In [69] a fault-tolerant fusion rule that employs error correcting codes to incorporate fault-tolerance capabilities is utilized for wireless sensor networks. Kapitanova *et al.* [70] propose a general failure detection, assessment, and adaptation approach devised for smart home applications. This approach utilizes semantics to deal with sensor faults which are often difficult to detect, and it could be in principle

leveraged for body sensor networks. The use of sensor fusion has been principally utilized in the AR domain to deal with sensor technological anomalies. Zappi *et al.* [51] show a significant tolerance increase by using a large set of sensors in combination with MV or bayesian fusion models. Multi-sensor fusion for counteracting the effects of calibration drift has been also utilized in [71]. Chavarriaga *et al.* [52] use an information theoretical approach to determine the level of noise in a sensor network. Concretely accuracy and diversity techniques are employed to select the most beneficial sensor nodes to compose the recognition ensemble. A similar model is presented in [53, 55] to detect anomalies and potential affected sensors in order to remove them from the sensor ecology. In a very recent work, Sagha *et al.* [56] have presented a more sophisticated model that not only attempts to detect the faulty sensors but also retrains the classifier models of the sensors detected as anomalous. In most of the previous approaches, MV is utilized for the ensemble fusion.

2.4. Sensor displacement in AR

Sensor displacement is a well-known problem in AR. Particularly critical for on-body inertial sensing, diverse approaches have been proposed in the literature to increase the robustness of AR methods to sensor displacements.

One main direction to deal with this issue consists in identifying displacement invariant features for recognition. Kunze *et al.* [72] studied how acceleration and gyroscope signals are affected by sensor displacement. They distinguished between gravitational, translational, and rotational components in the acceleration signal and showed that the acceleration component due to rotation is specially sensitive to sensor displacement. Based on this observation they proposed a heuristic method which achieved higher recognition rates for sensor displacement within a particular body part. Another approach that uses displacement invariant features was proposed by Foerster et al. [73]. By extracting signals from several locations within a body segment and applying a genetic algorithm for feature selection they identified features invariant to sensor displacement. They validated their approach using a human-computer interaction (HCI) and a fitness dataset and achieved improved recognition results with respect to standard features. These heuristic methods are coupled to the assumption that the user performs

the required specific activities at some point, which nevertheless might not be always guaranteed.

Another main direction to increase recognition robustness against sensor displacement is adapting the classifiers to the resulting shifts in the signal and feature distribution. Gao *et al.* [71] utilize an estimate of the constant gravity vector to transform the accelerometer signals from the device coordinate system, which is sensitive to the orientation, to the body reference coordinate system, which is seen to be steady. This estimation is essentially possible when the user remains static. The approach is evaluated for four categories of activities (lying down, sitting, standing and walking) demonstrating an increase of the overall accuracy when using the body reference frame with respect to the case in which the sensor original frame is considered. Also at the signal level, in [74] a robust motion direction of the user is obtained for sensors prone to rotation by using models based on principal component analysis. In [75, 76] the authors proposed an unsupervised adaptation method based on the expectation-maximization (EM) algorithm. They assumed that the anomalies introduced by sensor displacements can be characterized as a covariate shift [77]. They estimated this shift using an online version of the EM algorithm and transformed the sensor readings back in the feature space before classification. They tested their method on HCI, fitness and daily living scenarios. While the previous methods applied a transformation in the signal or feature space, in [78] the authors proposed an online self-calibration method to dynamically adapt the classifier model. The method consists of a calibration phase and a recognition phase. The calibration phase is triggered by the user when they observe that the recognition accuracy drops. In this phase the cluster centers of a nearest centroid classifier are adjusted at a predefined learning rate using the incoming instances after classification. The process is stopped when the gradient of the distance between the adapted class center and the original class center drops below a certain threshold. This method is intrusive and depends on the capacity of the user to identify erroneous recognition.

Multi-sensor fusion has been also utilized in the past for counteracting the effects of sensor orientation changes. In [51] the authors showed a significant tolerance increase by using a large set of sensors in combination with MV or NB decision fusion models. A more sophisticated scheme is presented in [53] which attempts to detect anomalies and potential affected sensors in order to remove them from the sensor ecology. This approach is further improved in [55] to also bring the system to

a new stable state. This is accomplished through a self-training process, which uses the fusion output to provide labels to retrain the sensor systems identified as anomalous. For all these works, the models are fundamentally validated on synthetically modeled rotations, but an important lack is observed with respect to the evaluation of sensor displacement due to translation.

2.5. Transfer learning in AR

Activity-recognition systems are usually designed around a set of sensors that are selected to be highly discriminative of the activities of interest [10, 79, 80, 81]. However, sensor configurations are prone to changes due to system maintenance and upgrades. Likewise, sensors availability may vary depending on the user context and needs during the course of a day. In fact, there is a multitude of sensors readily deployed by users themselves, with the widespread use of smartphones [82], and the rising availability of sensorized gadgets and smart-objects [83], or motion sensing shoes or garments [84], that may be arbitrarily accessed by users during their daily life. Living environments are also ever more richly sensorized in smart homes [85], but also chiefly for climate control, security, and ever more for entertainment systems supporting natural interaction [86].

Recently, diverse methods have been proposed to increase the tolerance to variability in the sensing environment at run-time by essentially substituting the sensor environment which was foreseen at design-time by the effective environment which is encountered at run-time. For example, sensor-placement-independent AR can be achieved by using datasets collected from multiple on-body locations [87]. This requires training data provided by the user. Self-calibration approaches require no user intervention, but were demonstrated only for specific cases (e.g. displacement of accelerometers [76, 78]). Combinations of multiple sensor modalities also help to tolerate on-body displacement [72], or to substitute sensor modalities [88]. However, these combinations must be predesigned for selected kind of variations. Alternatively, sensors can self-characterize their on-body placement [89] and orientation [90] to select the appropriate activity models at runtime, but this requires to predefine these models. One of the main limitations of most of these approaches is that they are basically constrained to foreseen run-time variations. Moreover, for each of them either activity models must be collected or ad-hoc transformations between modalities must

be designed. Transfer learning [91] is here devised to overcome these limitations, thus specially suitable to support unforeseen sensor setup run-time variations.

Practically, principles allowing a trained system to transfer AR capabilities to another system were proposed in [92, 93]. This approach works across sensor modalities and was characterized on different body-worn and ambient sensors. Essentially, an initial system recognizes activities and supervises the learning of a new one, without user specific intervention. Transfer learning methods were extended to operate on time series resulting from the activation of switches to transfer AR capabilities from one type of smart home to another kind of smart home with different and a-priori unknown number and placement of sensors [85]. These approaches operate on long time scales as they require all the relevant activities to be observed several times (e.g. timescale of days or more). Besides, these methods are prone to incomplete transfer learning since it is likely that the user does not perform the complete set of target activities. An exhaustive review and taxonomy of transfer learning approaches in AR is available in [3]. Here transfer methods are categorized along the data processing stage to which they apply, i.e., data acquisition and preprocessing, feature extraction, classification and eventually symbolic processing.

3

Tolerance of AR systems to sensor faults and failures

3.1. Introduction

Day by day we are witnessing the evolution of a society less tolerant to errors and mistakes. In this new technological era, users demand a seamless and fluent interaction with the technology of their daily living, and actually, they are already getting used to it. Annoying system outages, crashes, or reboots typical in the past are no longer tolerated, and could be the reason for a business collapse. In an increasingly competitive and diversified technological market, in which users may found hundreds or thousands of alternatives of a very product, important details such as the robustness of systems are compelling factors in the choice of one option instead of another. Moreover, people dissatisfaction not only translates into the rejection of a product that does not reach consumers expectations, but also may wage a smear campaign that could have a devastating impact on the brand image.

In AR applications, errors may come from how the sensor data stream is processed and/or be motivated by the sensor itself. Wearable sensor technology is subject to intensive use and potential harsh conditions, therefore prone to degradation and failures. For example, an instrumented bracelet may suffer from extreme temperature variations (e.g., from winter to summer), pressure (e.g., when carrying out objects or while leaning against a wall) or humidity (e.g., moisten or get soaked under heavy rain). Under these circumstances, sensors may degrade and the information delivered be corrupted or anomalous, therefore different to what is measured in normal conditions. At worst, sensors may completely fail or stop working and no data explicitly provided. Ignoring these potential anomalous situations may have a serious impact on system performance and usability, and by extension lead to a user's lack of interest in the product.

Classic AR approaches assume that the sensor configuration remains identical during the lifelong use of the system. Technological anomalies generally lead to changes in the sensor data streams, which are normally unforeseen during the design phase or unpredictable at runtime use. Consequently, models trained on ideal signal patterns may react in an undesired manner to imperfect or anomalous sensor data. This potentially translates into a partial or total malfunctioning of the AR systems. Given a malfunctioning situation, users normally need to resort to a maintenance service or expert, which entails new cost for the consumer, or the vendor as part of the warranty, apart from disturbing the continuity of use of the system and applications. For example, a sen-

sor may provide anomalous data because it is not adequately supplied, however, this situation could be reversed when the battery is recharged. For such a task it makes no sense to request the help of a support team, and for sure not throw the sensor out. AR systems should be able to notice the importance of the sensor anomaly and use the rest of the sensor ecosystem to overcome the effects of this issue. In the event of more critical or irrecoverable failures, assistance would be required but yet the recognition capabilities maintained until the system is fixed.

3.2. Technological anomalies

Electronic devices are subject to diverse type of technological or hardware anomalies. On-body inertial sensors are prone to changes in the bias, scale factors, non-linearity or electronic noise among others, normally due to decalibration or battery failures. Some of these anomalies have been extensively studied in the past, and actually important improvements have been made to minimize their effects. Nevertheless, some others are more difficult to handle from a hardware perspective and therefore more likely to appear during the use of these devices.

The principal effects of sensor hardware anomalies are depicted in Figure 3.1. Bias or offsets appear when the sensor internal characteristics such as the operation point alter due to environmental changes (e.g., ambient temperature), normally leading to variations in signal mean amplitude (i.e., DC component). Other typical artifacts are electronic noise and spurious spikes. Spurious spikes are short duration electrical transients often caused by ringing (oscillation of a voltage or current) or crosstalk (a circuit or channel induces an undesired effect in another circuit or channel). These electronic anomalies are normally due to parasitic capacitances, inductances or conductive coupling in the hardware circuit. These artifacts are not part of the design but just byproducts of the materials used to construct the circuit, however, they could be easily compensated through filtering and canceling techniques (e.g., high-pass filters for offset removal or low-pass filters for electronic noise cancellation). Nevertheless, other anomalies may imply a certain information loss that cannot be rectified through preprocessing techniques.

One of the most limiting data losing effects is associated to changes in the sensor dynamic range. These variations could appear due to a misconfiguration of the sensor or when the system is not adequately supplied. Sensor misconfigurations rarely take place, however, battery

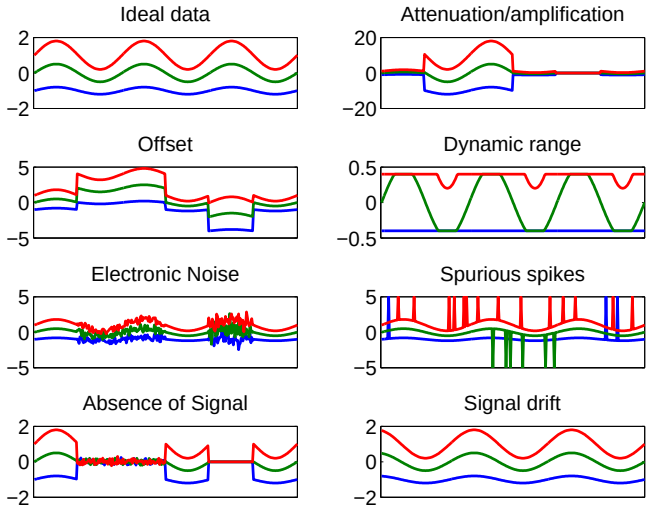


Figure 3.1: *Examples of the signal effects of the principal sensor technological anomalies. For the sake of interpretability, the anomalies are depicted for a set of basic signals.*

malfunctions may be more frequently observed. In the event of battery faults, it happens to occur that internal amplifiers and analog-to-digital converters are not appropriately fed and in consequence signal amplification and conversion incorrectly performed (new dynamic range). This translates into signal distortions such as flattening or skew, which basically correspond to an erroneous representation of the measured data, since these fall out of the bounds of the new dynamic range (see Figure 3.1, second row right column).

Another key shortcoming of on-body sensors, which applies to any wireless technology, is energy limitations. Sensor batteries are not of unlimited capacity and therefore need to be recharged from time to time. Moreover, sensor batteries lose charge with time and its capacity reduces as they are charged and drained. Therefore, there exist potential situations in which a sensor is not supplied, and consequently no data delivered. Permanent critical failures could appear in more extreme situations, for example when the sensor device falls to the ground, is accidentally hit or physically damaged in any other way. In these situations, not only could the sensor itself be broken but any other electronic

element necessary for the data acquisition and delivery destroyed (e.g., communication interface, processing unit), thus leading to an absence of signal or data (see Figure 3.1, bottom left corner).

Although some sensor anomalies may be solved during the ARC preprocessing phase (e.g., electronic noise and spurious spikes filtering), others are more difficult to overcome or avoid, specially at runtime. Under the event of a critical sensor failure or malfunctioning (e.g., sensor out of battery or totally broken), systems based on a single sensor unit demonstrate futile. Activity inference in SARC models rely on a single data stream, therefore no operation is logically possible since no data would be available. In these circumstances, the use of redundant or multi-sensor configurations appear to be a reasonable alternative. However, not all MARC models are seen to cope with the problem of a discharged or broken sensor. In fact, FFMARC models suffer from the same limitations as SARC to this respect. In FFMARC, features extracted from each sensor node are aggregated in a vector used as input to the classifier or reasoner, thus if data from a sensor is missing no features could be obtained for that node and thereby the feature vector is incomplete. In consequence, FFMARC approaches cannot operate and no activity-aware capabilities are then supported. Conversely, this problem is not seen to occur to DFMARC models. DFMARC approaches are based on the aggregation of the decisions computed from the individual processing of each sensor data stream. Therefore, even whether a sensor data stream could not be available, a decision may be made from the combination of the decisions obtained from the remaining active sensors.

HD and MV were introduced in Section 2.2.4 as the two principal models used for DFMARC. However, both HD and MV models demonstrate significant weaknesses when dealing with sensor technological anomalies. An example is used to illustrate this. Let us consider a sensor setup consisting of seven sensors worn on diverse body parts. For a given AR problem, the models are trained and performance metrics obtained. From this, individual sensor performances[1] (e.g., accuracy in %) could be characterized for each respective sensor S_i, e.g., S_1=99%, S_2=85%, S_3=82%, S_4=55%, S_5=39%, S_6=36%, and S_7=31%. Now,

[1]These performances could represent the recognition capabilities of the systems for a given activity problem, but may differ for another problem. This justifies the use of sensors that may not be found much reliable for the original target problem but be of much utility for the discrimination of other activities of interest for other applications.

when the HD model is employed, the system accuracy will principally depend on the performance of the decisor ranked on top of the hierarchy (i.e., S_1). If a low-ranked sensor (e.g., $S_5 - S_7$) gets unavailable, no variation is expected on the overall system performance. However, if sensor issues affect the top-ranked decisors (e.g., S_1, S_2) the recognition capacity may get seriously deteriorated. For example, if S_1 gets out of battery the bulk of the decisions would rely on S_2, thus a drop on the recognition performance is expected. Differently to HD, MV actually leverages the disappearance of sensors that yield poor decisions. Thus for this example, if S_5, S_6 or S_7 is shutdown, it could be anticipated an improvement of the recognition performance (in fact the more low-rated sensors get unavailable the better the performance will in principle be). Unfortunately, this also applies the other way around. Then, if a high-rated sensor becomes unavailable, the chances of misrecognition increases, specially when this results in a majority of low-rated decision makers. For example, if S_1 gets out of the pool of decision makers, a plurality of low-rated sensors (S_5, S_6, S_7) overtakes a minority of higher-rated sensors (S_2, S_3) and the erratic decisions provided by S_4, thus leading to an overall low performance. According to this, HD and MV models are qualitatively shown to have limited capabilities to deal with sensor failures.

In the previous example, the most beneficial approach would consist in relying the activity detection on the decisions made by the highest-rated sensors (i.e., S_2 and S_3), which roughly corresponds to the combination of HD and MV models. In next section, a novel DFMARC model that takes in this idea and leverages the potential of the remainder of active sensors is presented.

3.3. Hierarchical weighted classifier

Taking into account the advantages and drawbacks of HD and MV approaches, a new ensemble model is here presented to cope with the effects of sensor anomalies. The model combines the decisions provided by each sensor entity, making them all participatory on the decision process, but also ranking the relative importance of each decision through the use of weights based on the individual performance of each classification unit. Moreover, decisions are not only combined at the sensor level but also at the activity level, which is devised to increase recognition systems reliability and robustness, and specially important to support flexible AR setups (Section 3.4.6).

The proposed model [94], hereafter called hierarchical weighted classifier (HWC), is composed by three decision making levels or stages (see Figures 3.3-3.4). Given a scenario with M nodes of information (sensors) and N classes (activities), a set of M by N base classifiers (c_{mn}, $\forall\ m = 1, \ldots, M, n = 1, \ldots, N$) is defined. These are binary classifiers specialized in the discrimination of the activity or class n by using the information obtained from the sensor or node m. Each base classifier applies an one-*vs*-rest binary classification strategy[2], which further allows for the use of any type of standard classification paradigm. This defines the first layer of the model, here identified as base, class or *activity level*. The second classification level, here *sensor level*, is defined through M node or sensor classifiers (S_m, $\forall\ m = 1, \ldots, M$). Sensor classifiers are not machine learning-type entities, but decision making structures. Each sensor classifier is composed by N base classifiers (one per class), whose decisions are combined through an activity-dependent weighting scheme. Finally, the last layer, here called *network level*, is in charge of the weighted aggregation of the decisions given by each sensor classifier, eventually providing the recognized activity or class. The weights used in the network level depends on the recognition capabilities of each individual sensor classifier.

The model training requires just a few steps (see Figure 3.2). Firstly, the training dataset is partitioned into three equally-sample-distributed parts to approximately cover the same sample space. One of these partitions is used to train all base classifier entities. After training, a second partition is used to test the performance of each base classifier. From here, statistical metrics are obtained and further used to define the first level of weighting parameters (α, β). Sensor classifiers are completely defined at this point. Then, the weighting parameters for the network level are assessed. To that end, the third yet unused part of the dataset is considered now for the evaluation of each individual sensor classifier. From their performance statistics the diverse network level weights are obtained (γ, δ). The HWC is at this point completely characterized.

For this structure two weighting schemes are proposed. The first approach (Figure 3.3) consists in the weighting of the decisions given by each classifier by using a single weight that emphasizes the contributions of insertions and rejections[3] in a similar way (i.e., *unified*

[2]Other approaches as the one-*vs*-one may be similarly applied but here the one-*vs*-rest is particularly recommended to minimize the number of classification entities.

[3]In machine learning [95], insertions (hits) and rejections (deletions) respectively

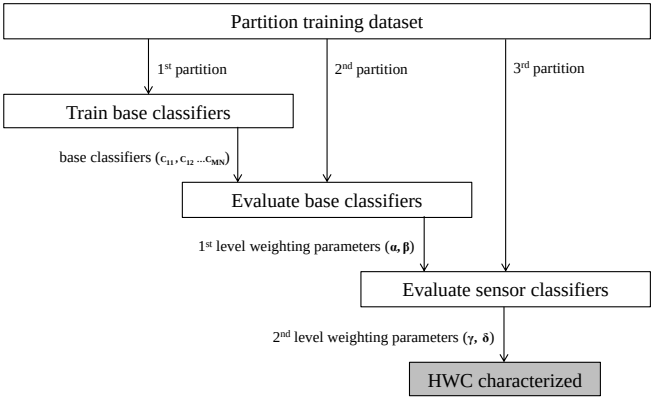


Figure 3.2: *Training steps of the HWC model.*

weighting). The second approach (Figure 3.4) corresponds to an improved version in which two independent weights are used to respectively weight insertions and rejections (i.e., *insertion-rejection weighting*). In the following, the process to compute the weights and decision aggregation models are described for both approaches.

Unified weighting model (alfa-gamma)

The model characterization process starts by evaluating the individual average accuracy of each base classifier. Since these are binary classifiers the accuracy is computed as follows (see Section 2.1.7):

$$\overline{R_{mn}} = \frac{TP_{mn} + TN_{mn}}{TP_{mn} + FP_{mn} + FN_{mn} + TN_{mn}} \tag{3.1}$$

with TP_{mn} (true positives), FP_{mn} (false positives), TN_{mn} (true negatives) and FN_{mn} (false negatives) computed from the evaluation of the classifier c_{mn}. This process is repeated for each individual node. From these figures, the base classifier weights may be obtained through averaging across all values:

refer to positive and negative classifications. For the one-vs-rest case, an insertion is observed when the classifier recognizes a class as belonging to its class of specialization, while a rejection is produced when the class is identified as one of the rest of classes.

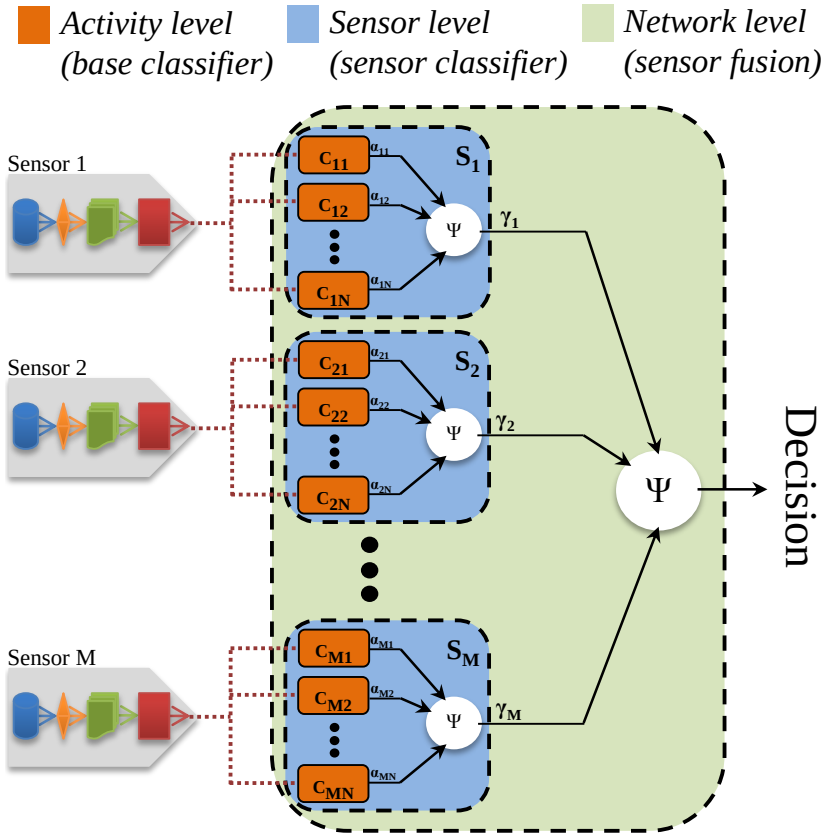


Figure 3.3: *Structure of the HWC for the unified weighting scheme ($HWC_{\alpha\gamma}$). The features extracted from each sensor are used as inputs to a set of N by M base classifiers (c_{mn}). Classifiers insertions are α-weighted while rejections are ignored. These decisions are aggregated through a combiner function (Ψ), thus yielding a decision for each sensor classifier (S_m). The decisions made across all sensor classifiers are γ-weighted and once again combined to provide the eventual recognized activity.*

$$\alpha_{mn} = \frac{\overline{R_{mn}}}{\sum_{k=1}^{N} \overline{R_{mk}}} \tag{3.2}$$

These weights represent the importance that each base classifier will have on the sensor classifier decision scheme. A specific voting algorithm is considered to fuse all base classifiers decisions for each corresponding sensor classifier. For a sensor m, given a window instance s_{mk} characterized through the corresponding feature vector $f_m(s_{mk})$, and being q the activity or class predicted by c_{mn} for that instance, if such class belongs to the class of specialization of c_{mn} (i.e., $q = n$) this classifier will set its decision to α_{mn} for the class n and 0 for the rest of classes. The opposite is made for $q \neq n$. This could be written as ($\forall\ \{q, n\} = 1, \ldots, N$):

$$WD_{mn}\left(f_m(s_{mk})\right) = \begin{cases} \alpha_{mn}, & f_m(s_{mk}) \text{ classified as } q \\ 0, & f_m(s_{mk}) \text{ not classified as } q \end{cases} \quad (\forall\ q = n)$$
$$\begin{cases} \alpha_{mn}, & f_m(s_{mk}) \text{ not classified as } q \\ 0, & f_m(s_{mk}) \text{ classified as } q \end{cases} \quad (\forall\ q \neq n) \tag{3.3}$$

where WD_{mn} represents the weighted decisions of the classifier c_{mn}. To compute the output of the *m-th* sensor classifier (O_m), the weighted decisions of each classifier are aggregated through a combiner function (ψ). Here, a cumulative sum function is specifically used across all classes:

$$O_{mq}\left(f_m(s_{mk})\right) = \sum_{n=1}^{N} WD_{mn}\left(f_m(s_{mk})\right) \tag{3.4}$$

The class predicted by S_m is the class q for which the sensor classifier output is maximized:

$$q_m\left(f_m(s_{mk})\right) = \underset{q}{\operatorname{argmax}}\left(O_{mq}\left(f_m(s_{mk})\right)\right) \tag{3.5}$$

At this stage, sensor level classifiers are fully defined. In fact, each sensor classifier could be already used in a SARC mode. Nevertheless, the HWC is rather devised for MARC approaches.

Similarly to as performed for Equation 3.2 a weight γ_m is now calculated for weighting the decisions provided by each sensor classifier

S_m. This corresponds to the weights of the network level. To do so, the accuracy performance rate of each sensor classifier is evaluated:

$$\overline{R_m} = \frac{\sum_{i=1}^{N} tp_i}{Q_m} \tag{3.6}$$

with tp_i the diagonal elements and Q_m the cumulative sum of all the elements of the confusion matrix obtained from the evaluation of S_m (see Figure 2.3). Averaging across all sensor classifiers, γ_m is computed as follows:

$$\gamma_m = \frac{\overline{R_m}}{\sum_{k=1}^{M} \overline{R_k}} \tag{3.7}$$

The output at the network level is now calculated taking into account the individual outputs obtained from each sensor classifier. Given a sample s_k defined through the corresponding data windows obtained from each respective sensor ($\{s_{1k}, s_{2k}, ..., s_{Mk}\}$), and being characterized through their corresponding feature vectors ($\{f_1(s_{1k}), f_2(s_{2k}), ..., f_M(s_{Mk})\}$), the aggregated output is:

$$O_q\left(f(s_k)\right) = O_q\left(\{f_1(s_{1k}), f_2(s_{2k}), ..., f_M(s_{Mk})\}\right) = \sum_{p=1}^{M} \alpha_p O_{pq}\left(f_p(s_{pk})\right) \tag{3.8}$$

Similar to Equation 3.5 the eventually yielded class q is obtained as:

$$q = \underset{q}{\operatorname{argmax}}\left(O_q\left(f(s_k)\right)\right) \tag{3.9}$$

At this point the HWC is simply defined through the trained class classifiers (c_{mn}), the class level weights (α_{mn}) and the sensor level weights (γ_m).

Insertion-rejection weighting model (alfa-beta-gamma-delta)

The main difference of this second model with respect to the previous one is that insertions and rejections are independently weighted. This approach may leverage the potential of classifiers that may be accurate inserters or rather be better qualified as rejecters, or both. For

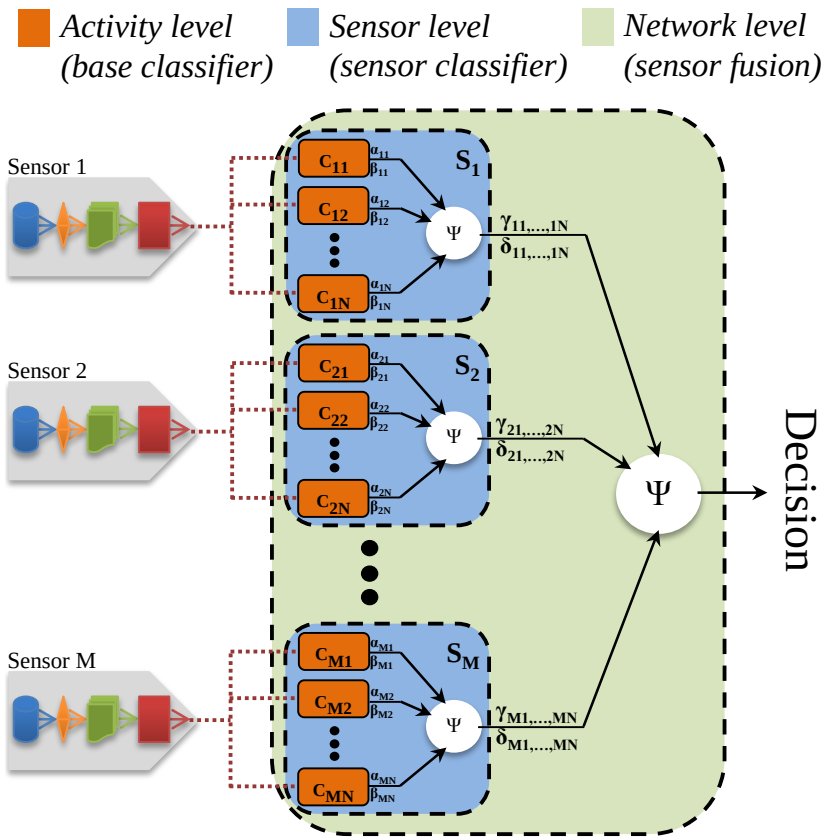


Figure 3.4: *Structure of the HWC for the insertion-rejection weighting model ($HWC_{\alpha\beta\gamma\delta}$). The features extracted from each sensor are used as inputs to a set of N by M base classifiers (c_{mn}). Classifiers insertions and rejections are respectively α- and β-weighted. These decisions are aggregated through a combiner function (Ψ), thus yielding a set of decisions for each sensor classifier (S_m). The decisions made across all sensor classifiers are γ-weighted (insertions) and δ-weighted (rejections), and once again combined to provide the eventual recognized activity.*

example, a classifier c_{mn} could be very precise when detecting data windows that belongs to its class of specialization ($q = n$), but fail when

distinguishing other classes from this ($q \neq n$). On the other hand, a base classifier may be very accurate when identifying data windows as not belonging to its specialization class ($q \neq n$) but imprecise when detecting the actual class of specialization ($q = n$).

The process followed to characterize the insertion-rejection weighting model is quite similar as for the unified model. At the activity level two weights are obtained. These parameters are defined as α_{mn} and β_{mn}, and respectively represent the insertion and rejection weights for c_{mn}. The values of α_{mn} and β_{mn} are obtained from the performance assessment of c_{mn}. In particular, α_{mn} corresponds to the sensitivity whilst β_{mn} relates to the classifier specificity. These performance parameters have been selected since these represent well insertion and rejection capabilities of the classifier. As described in Section 2.1.7, given TP_{mn} (true positives) the number of correctly identified samples, FP_{mn} (false positives) the incorrectly identified samples, TN_{mn} (true negatives) the number of correctly rejected samples and FN_{mn} (false negatives) the incorrectly rejected samples, all specifically computed from the evaluation of the classifier c_{mn}, α_{mn} and β_{mn} define as:

$$\alpha_{mn} = \frac{TP_{mn}}{TP_{mn} + FN_{mn}} \tag{3.10}$$

$$\beta_{mn} = \frac{TN_{mn}}{TN_{mn} + FP_{mn}} \tag{3.11}$$

Akin to the unified model, a voting method is considered to fuse all base classifiers decisions for each corresponding sensor classifier. For a sensor m, given a window instance s_{mk} characterized through the corresponding feature vector $f_m(s_{mk})$, and being q the activity or class predicted by c_{mn} for that instance, if such class belongs to the c_{mn} class of specialization ($q = n$) this classifier will set its decision to α_{mn} for the class n and to 0 for the rest of classes. Otherwise ($q \neq n$), the classifier decision is set to 0 for the class n and to β_{mn} for the others. In summary, the classifier c_{mn} weighted decision (WD_{mn}) for the class q may be defined as ($\forall\ \{q, n\} = 1, \ldots, N$):

$$WD_{mn}\left(f_m(s_{mk})\right) = \begin{cases} \alpha_{mn}, & f_m(s_{mk}) \text{ classified as } q \\ 0, & f_m(s_{mk}) \text{ not classified as } q \end{cases} (\forall\, q = n)$$
$$\begin{cases} \beta_{mn}, & f_m(s_{mk}) \text{ not classified as } q \\ 0, & f_m(s_{mk}) \text{ classified as } q \end{cases} (\forall\, q \neq n) \tag{3.12}$$

The aggregation of the weighted decisions provided by each base classifier for the *m-th* sensor classifier (S_m) may be computed as follows:

$$O_{mq}\left(f_m(s_{mk})\right) = \sum_{n=1}^{N} WD_{mn}\left(f_m(s_{mk})\right) \tag{3.13}$$

The class predicted by S_m is the class q for which the sensor classifier output is maximized:

$$q_m\left(f_m(s_{mk})\right) = \underset{q}{\operatorname{argmax}}\left(O_{mq}\left(f_m(s_{mk})\right)\right) \tag{3.14}$$

For the next level, similar parameters to α_{mn} and β_{mn} are obtained, here defined as γ_m (insertions) and δ_m (rejections). Nonetheless, the way these are computed varies slightly with respect to the formers. At the network level the classifiers are not binary but multiclass models. Therefore, the evaluation of each sensor classifier requires to extend sensitivity and specificity concepts to the multiclass case (see details in [96]). According to this generalization, γ_m and δ_m may be described as:

$$\gamma_m = \langle \gamma_{m1}, \gamma_{m2}, \ldots, \gamma_{mn} \rangle = \left\langle \frac{TP_{m1}}{TP_{m1} + FN_{m1}}, \frac{TP_{m2}}{TP_{m2} + FN_{m2}}, \ldots, \frac{TP_{mn}}{TP_{mn} + FN_{mn}} \right\rangle \tag{3.15}$$

$$\delta_m = \langle \delta_{m1}, \delta_{m2}, \ldots, \delta_{mn} \rangle = \left\langle \frac{TN_{m1}}{TN_{m1} + FP_{m1}}, \frac{TN_{m2}}{TN_{m2} + FP_{m2}}, \ldots, \frac{TN_{mn}}{TN_{mn} + FP_{mn}} \right\rangle \tag{3.16}$$

where $\{TP/TN/FP/FN\}_{mn}$ refer to the classification counting values, but here computed for each class k across the confusion matrix results obtained from the evaluation of S_m ($\forall\, m = 1, \ldots, M, n = 1, \ldots, N$).

Since decisions are made in a multiclass fashion, γ_m and δ_m are used to reward or penalize each considered class. Accordingly, given q_m the decision of S_m for the sample s_{mk}, the set of weighted decisions from this classifier is defined as:

$$WD_m\left(q_m\left(f_m(s_{mk})\right)\right) = \begin{cases} \gamma_{mn}, & n = q_m\left(f_m(s_{mk})\right) \\ -\delta_{mn}, & n \neq q_m\left(f_m(s_{mk})\right) \end{cases} \quad (\forall\, n = 1, \ldots, N) \tag{3.17}$$

The output at the network level is now calculated taking into account the individual outputs obtained from each sensor classifier. Given a sample s_k defined through the corresponding data windows obtained from each respective sensor $(\{s_{1k}, s_{2k}, ..., s_{Mk}\})$, and being characterized through their corresponding feature vectors $(\{f_1(s_{1k}), f_2(s_{2k}), ..., f_M(s_{Mk})\})$, the aggregated output is:

$$O_q\left(f(s_k)\right) = O_q\left(\{f_1(s_{1k}), f_2(s_{2k}), ..., f_M(s_{Mk})\}\right) = \sum_{p=1}^{M} WD_p\left(q_p\left(f_p(s_{pk})\right)\right) \tag{3.18}$$

Finally, the class q yielded on top of the hierarchy is obtained as:

$$q = \underset{q}{\operatorname{argmax}}\left(O_q\left(f(s_k)\right)\right) \tag{3.19}$$

3.4. Evaluation of AR systems tolerance to sensor technological anomalies

In Section 3.2 it was qualitatively shown that some of the most widely used AR approaches are not capable of operating under the event of critical hardware anomalies. Moreover, models that could in principle be thought more robust to sensor faults were also demonstrated of limited tolerance as illustrated for a particular example. In Section 3.3, the HWC model was presented as an alternate new approach to cope with the effects of sensor technological. This section aims at quantitatively demonstrating the utility of this approach. To that end, the HWC model is first compared to standard AR models in ideal conditions (Section 3.4.3), in order to prove it provides similar recognition capabilities

to them. Next, the model is evaluated for the case in which critical sensor failures are assumed, i.e., sensor do not deliver data (Section 3.4.4). Finally, the recognition capabilities under moderate sensor faults are also assessed (Section 3.4.5). The experimental evaluation is performed on a well-known dataset broadly used in the wearable computing domain. This dataset is described in the following (Section 3.4.1).

3.4.1. Benchmark dataset

Sensor technological anomalies normally appear in a random and occasional manner, thereby it could be complicated to find them during experimental recordings. Nevertheless, an interesting characteristic of these anomalies is that their effects may be reasonably easy modeled. Therefore, the approach followed in this work consists in synthetically introducing sensor hardware anomalies on the activity data experimentally recorded in a daily living setting.

The activity dataset used in this work was first introduced in [10], and has been repeatedly used to benchmark AR models [97], something which is quite important in a field that lacks of specific gold standards. This dataset has been considered specially interesting for evaluation since not only includes data collected in a natural out-of-lab settings but for a diverse sample population and activities. Moreover, this is one of the few datasets that is publicly available[4].

Training and evaluation of previous AR systems have been normally performed on data collected in laboratory settings. However, when these very systems are used in realistic scenarios, their performance may severely worsen. This is normally consequence of the artificial constriction, simplification or influence that laboratory environments may induce in people normal behavior. These data do not normally capture relevant facets of a person's daily life such as interaction with their environment (objects, people, etc.) or absence of self-awareness when carrying out habitual activities. To address this problem, two scenarios are explored in this dataset, one corresponding to a classical in-lab setting and a second one on which a semi-naturalistic scenario is defined. For the semi-naturalistic scenario, users are succinctly oriented to perform an activity that is not the actual action of interest. For example, instead of asking a user to sit down and relax, the user is asked to watch TV during a certain period of time on which "sitting"

[4]Dataset files and description could be obtained at http://architecture.mit.edu/house_n/data/Accelerometer/BaoIntilleData04.htm

is recorded. Other examples are to stare in front of a painting (to record "standing still"), sleep/rest (to record "lying down") or move from one stage to another (to record "climbing stairs"). To suppress any type of influence from the presence of experts during the recordings, supervision or observation of researchers is avoided. A worksheet is provided instead to each participant with the instructions of the course to perform. Participants are asked to execute a set of everyday tasks such as "walking", "folding laundry" or "riding escalator", but no hint or special indication on how these activities must be performed is given by experts.

People's activities are monitored through the use of wearable inertial sensors devised to capture human body motion. Concretely, five bi-axial accelerometers[5] are employed to register the motion experienced by the subjects' right hip, dominant wrist, non-dominant arm, dominant ankle and non-dominant thigh respectively. Twenty subjects aged 17 to 48 participated in the study in two runs each (laboratory/semi-naturalistic). The data are registered at a sampling rate of 76.25Hz. Multi-sensor data synchronization is achieved through the use of quartz clocks and sinusoidal patterns recorded at the start and end of each monitoring session. These patterns are obtained through simultaneous shaking of all sensors in a given direction. For more specific details on the hardware used, protocol and data statistics the reader is referred to [10].

From the complete activity set, the most representative nine activities were selected (Figure 3.5), covering from intense activities as running or cycling to fitness exercises as stretching, or sedentary activities as sitting or lying down. The decision of leaving out some of the original activity set is motivated by various reasons. During the data inspection, missing or erroneously labeled data were encountered for some activities and subjects. The use of these data may drive to an unbalanced and possibly unfair comparison of the capabilities of the systems when discriminating among the diverse activities. Others, although relevant were found redundant, such as "sitting and relaxing", "sitting and working on computer" and "sitting and watching TV", thus defining some isolated clusters of activities. Finally, previous work demonstrates that for the detection of some particular actions the com-

[5]Although the unit of measure of acceleration in the International System of Units (SI) is m/s^2, in most applications these are normally referred to the 'g'-value, which approximately corresponds to $9.78 m/s^2$. For convenience, this representation is used in this work.

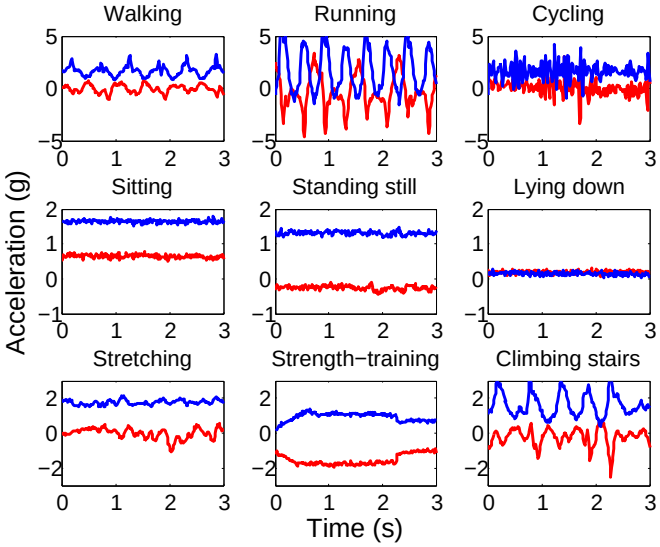


Figure 3.5: *Snapshots from the nine activities selected for evaluation. Data correspond to acceleration signals (red, X-axis and blue, Y-axis) registered through the arm sensor and for a particular subject.*

#Subjects	Age	#Activities	Sensors placement
20	*17-48*	*9*	*H = Hip, W = Wrist, A = Arm, K = Ankle, T = Thigh*

Table 3.1: *Experimental dataset description summary.*

bined use of on-body sensors with other sensing modalities is better recommended (e.g., RFID sensors for kitchen tasks [98]). In either case, this number of activities is found representative for the purpose of this study, especially taking into account the average number of activities considered in previous related work. A summary of the experimental dataset used in this work is presented in Table 3.1.

3.4.2. Experimental setup

For the activity recognition process, the diverse stages of the ARC are implemented[6] (see Section 2.1). Raw acceleration signals are ac-

[6]All the processing is performed in Matlab R2011b. For the preprocessing, featuring and classification stages some of the functions provided in the *Signal Pro-*

quired through the on-body inertial sensors. The recorded signals are affected by spurious spikes and electronic noise, sensor anomalies that can be eliminated during the signal preprocessing phase. To remove these anomalies a 20Hz cutoff low pass elliptic FIR filter is used. This is demonstrated not to eliminate valid information for daily physical activity assessment [99, 100].

The signals are subsequently partitioned into windows of data of approximately 6 seconds as suggested in [10]. Next, features are extracted to characterize each window data. To analyze the required classification complexity for this problem, diverse feature vector lengths are tested (1, 5, 10 and 20 features respectively). Here a subset of the complete group of features proposed in a previous work is considered [101]. These features corresponds to a combination of statistical functions such as median, kurtosis, mode, range, and magnitudes or functions obtained from a domain transformation of the original data such as energy spectral density, spectral coherence or wavelet coefficients ("a1 to a5" and "d1 to d5" Daubechies levels of decomposition) among others. The best features ranked through the use of a receiver operating characteristic feature selector [95] are chosen until complete the feature vector lengths defined for each case. Specific of each ARC model (Section 2.2), SARC builds on a single feature vector extracted from the data of the corresponding sensor used, FFMARC aggregates all feature vectors computed across the five sensor streams into a single feature vector, and DFMARC uses the feature vector extracted from each sensor as input to each respective node classifier.

The classification stage is also different for each ARC model. SARC and FFMARC employ standard multi-class classifiers such as the presented in Section 2.1. Concretely, a C4.5 implementation [102] is used for DT, the approach presented in [95] for NB and empirically k-tuned KNN models similar to the described in [103]. Also SVM models are used for comparison, here implementing a radial basis function (RBF) kernel with automatically tuned (grid search) hyper-parameters γ and C [104]. These standard classifiers are also used as core of the base classifiers used in the DFMARC approaches (here, HD, MV and HWC).

The evaluation of the systems is carried out through a ten-fold random-partitioning cross-validation to support an adequate comparison among the models (Section 2.1.7). Moreover, this process is repeated 100 times to ensure statistical robustness and procure the con-

cessing, *Statistics* and *Bioinformatics* toolboxes have been used, while many others have been specifically defined for this purpose.

vergence to a correct estimation of the performance of the systems. In each iteration the data sample is arbitrarily partitioned for the cross-validation process.

Since this work aims at improving the understanding of the operation of AR systems in real-world conditions, the data collected for the semi-naturalistic scenario is particularly employed for evaluation.

3.4.3. Performance in ideal conditions

The HWC is here particularly proposed to deal with the effects of sensor failures and degradations, however, not only should the model be useful to overcome these limitations but to be valid for AR tasks in normal circumstances. Therefore, it is found necessary to first evaluate its recognition capabilities in invariant setups. Moreover, the HWC performance is compared to the baseline determined by standard AR systems.

The HWC was defined from the observation of HD and MV limitations to sensor failures. In fact, the HWC combines the main advantages of both models, therefore it seems reasonable to first compare the recognition performance for the three of them. In Figure 3.6, the confusion matrices computed from the performance assessment of HD, MV and HWC models for diverse feature vectors are shown. For the HWC, the two proposed weighting models (unified weighting, $\text{HWC}_{\alpha\gamma}$; insertion-rejection weighting, $\text{HWC}_{\alpha\beta\gamma\delta}$) are respectively evaluated. From the results, the HWC clearly exceeds the recognition performance shown by HD and MV models. Moreover, this happens to occur for all classification paradigms and independently of the complexity of the feature vector used. In fact, promising results are already obtained for the case in which a sole feature is used for each base classifier, especially for the $\text{HWC}_{\alpha\beta\gamma\delta}$ approach. The performance proves to be close to absolute (confusion matrices almost diagonal) when richer feature vectors are used.

HD and MV models show worse recognition capabilities. A severe misclassification rate is observed when simple feature sets are employed. The performance is nevertheless enhanced when 10 and 20 features are used (Figures 3.6(c)-3.6(d)), specially for the HD model. This is motivated by an improvement of the recognition capabilities of each individual node or sensor classifier. Then, HD models yield more accurate decisions on top of the hierarchy and consequently reduce the errors made through this. Decisions computed through MV are easily

corrupted when the number of low performance decision entities overtake the accurate ones. In such circumstances, the potential of high performance sensor classifiers is hidden by a plurality of less accurate node classifiers, introducing a systematic error that degrades the performance of the whole recognition system. To compensate this situation, a majority of accurate sensor classifiers is required, which is here encountered as the number of features increases.

From the previous results, it could be also concluded that the $\text{HWC}_{\alpha\beta\gamma\delta}$ model surpasses the $\text{HWC}_{\alpha\gamma}$ approach. Actually, both models present a very high performance, which demonstrates the potential of the HWC structure, however, the slight improvement shown for the $\text{HWC}_{\alpha\beta\gamma\delta}$ approach also proves the importance of the weighting scheme. For the sake of simplicity and since it proves to be the most accurate approach, the $\text{HWC}_{\alpha\beta\gamma\delta}$ will be used in advance as predominant HWC model.

Most of the contributions in the AR domain are based on a SARC or FFMARC model. These approaches have demonstrated good recognition capabilities in a wide sort of AR problems. In this way, SARC and FFMARC models could be used to define a performance baseline for recognition in ideal circumstances. For the here considered experimental setup, five SARC models are devised (one per sensor) whilst the fusion of the features extracted from each sensor data stream is implemented for the FFMARC approach. In Figure 3.7, the performance results obtained from the evaluation of SARC, FFMARC and HWC models are depicted. From here, FFMARC and HWC clearly outstand as the most accurate models. In general, the FFMARC outperforms the HWC for simple feature sets (1 and 5 features). This is quite reasonable since the aggregated feature vector used in FFMARC is richer than the used for each HWC base classifier. This is especially relevant when a few features are computed. For example, when a single feature is extracted from each sensor stream, base classifiers operate on a 1-dimensional (1-D) feature space while a 5-D space is used to characterize the activity data in the FFMARC. Anyway, the differences are not higher than 7% accuracy at worst case. The gap between both models reduces as more features are employed, with the HWC the most accurate approach for some cases, and performance levels that represent an almost perfect discrimination of the activities.

SARC models provide discrete results, specially for reduced feature sets. As it occurs for the rest of the models, higher performance values are obtained as the number of features increases. This proves to be

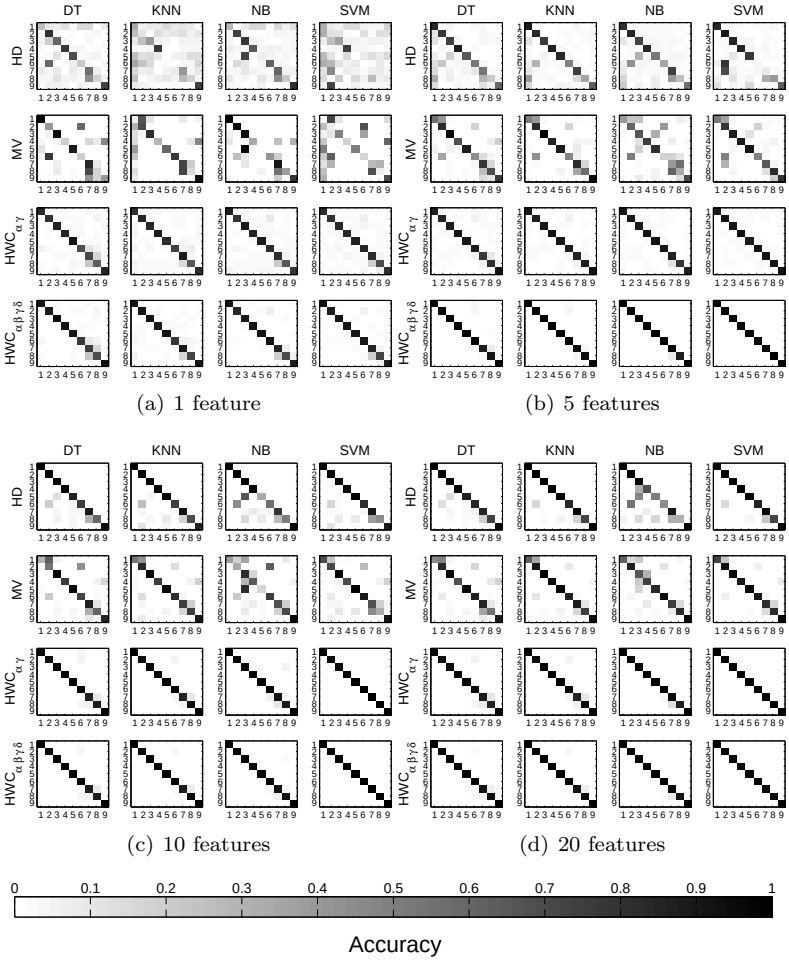


Figure 3.6: *Confusion matrices obtained from the experimental evaluation of each DFMARC modality (HD, MV, $HWC_{\alpha\gamma}$, and $HWC_{\alpha\beta\gamma\delta}$) and machine learning paradigm (DT, KNN, NB, and SVM). Diverse feature vector lengths are considered (**(a)** 1, **(b)** 5, **(c)** 10, and **(d)** 20 features). Confusion matrix legend (activities): 1 = Walking, 2 = Running, 3 = Cycling, 4 = Sitting, 5 = Standing, 6 = Lying down, 7 = Stretching, 8 = Strength-training, 9 = Climbing stairs.*

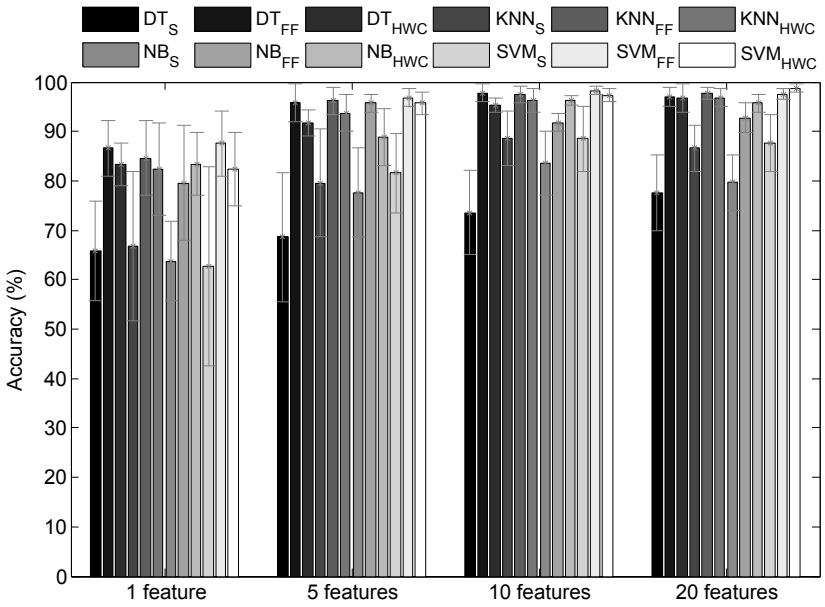


Figure 3.7: *Accuracy (average - bar - and standard deviation - whiskers -) results from the evaluation of the SARC (S), FFMARC (FF) and HWC (HWC) approaches. Results are averaged across all sensors for the SARC model. The insertion-rejection weighting scheme is particularly used for the HWC model. Diverse feature vector lengths are considered for evaluation (**(a)** 1, **(b)** 5, **(c)** 10, and **(d)** 20 features). Legend: <classification paradigm>$_{<AR\ approach>}$.*

very important for this model, whose best results are obtained for the 10 and 20 features case, but yet below 90% accuracy. This increase in the recognition capabilities of single sensor systems, which are directly combined in HD and MV, also explains the similar behavior described for these DFMARC models.

To put it in a nutshell, from the previous results the HWC shows as reliable as standard AR approaches in ideal circumstances, however, in theory it further surpasses all the previously tested standard models in terms of tolerance to sensor technological anomalies. This is investigated in the next two sections.

3.4.4. Tolerance to sensor failures

In Section 3.2, sensor technological anomalies were categorized in critical sensor errors or failures that suppose a complete data loss, and sensor faults which translates into a degeneration of the sensor data with respect to ideal conditions. From these, sensor failures pose the worst changes in the sensor setup. A sensor may fail to deliver data because its battery is depleted or an electronic component broken. Similar effects are practically seen when a sensor is left behind or shutdown. Therefore, the study presented in this section not only applies to critical sensor failures but to any circumstance that represent a sensor disappearance or removal from the original or default setup.

In the event of a sensor failure or shutdown, SARC and FFMARC models are shown not to operate. Since no data are provided, SARC models lack of practical information to proceed. FFMARC may utilize the information of the remaining active sensors, however, the aggregated feature vector cannot be in principle built since the values from the affected sensor are missing. To keep the system operating, it could be implemented a mechanism to replace the sensor missing values by artificially generated data. Nevertheless, introducing artificial data may have important consequences since this information may correspond to data measured in normal conditions (e.g., missing values filled with '0s' could represent actual values measured for some activities). Moreover, this turns to be less and less viable as the number of unavailable sensors increases.

Conversely to SARC and FFMARC models, the HWC model was devised to face the situation of having missing sensors. In the following, the recognition capabilities under the event of critical sensor failures are analyzed. To that end, a HWC model designed for the ideal or predefined sensor setup (i.e., the five sensors: hip -H-, wrist -W-, arm -A-, ankle -K-, and thigh -T-) is tested on diverse setup configurations respectively corresponding to cases in which data from a sensor or multiple sensors are not available. Concretely, the $\text{HWC}_{\alpha\beta\gamma\delta}$ with KNN and ten features is employed given its demonstrated high quality for recognition in ideal conditions (Figures 3.6-3.7). The results of this evaluation are summarized in Figure 3.8.

For the case in which one sensor stops delivering data the performance is observed to practically remain identical to when the complete set of sensors is used (>97% accuracy). A subtle performance worsening ($\leq$2%) is only seen for the case in which T or W sensors are

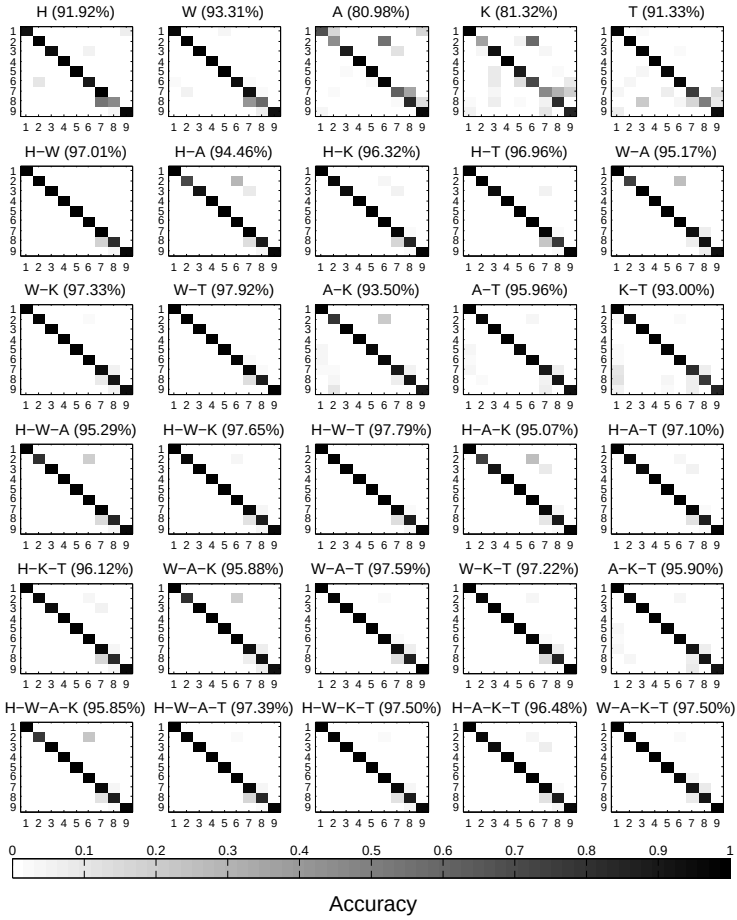


Figure 3.8: *Confusion matrices for the $HWC_{\alpha\beta\gamma\delta}$ model for all possible sensor setup configurations after the effect of sensor failures. KNN and the ten features setting is used for the base classifiers. Top title of each confusion matrix identifies active sensors and overall accuracy (in brackets). Sensors legend: H = Hip, W = Wrist, A = Arm, K = Ankle, T = Thigh. Confusion matrix legend (activities): 1 = Walking, 2 = Running, 3 = Cycling, 4 = Sitting, 5 = Standing, 6 = Lying down, 7 = Stretching, 8 = Strength-training, 9 = Climbing stairs.*

missing. A remarkable tolerance to the failure of two sensors is also observed. For some sensor configurations the performance is in the range of what is obtained for the default setup, at worst dropping the performance a bit more than for the case of one unavailable sensor. Not only may the HWC deal with failures on two sensors but even on three of these. In fact, the performance remains unaltered for some combinations of active sensors (W-T, W-K, H-W), and almost no worsening is seen but for a few cases. Yet, the performance drop is lower to 4% for those cases. At worst, the AR system must rely on the data captured through a single active sensor. Under that circumstances, the recognition capabilities worsen, although differently for each sensor type. The system accuracy is superior to 91%, this is a drop of less than 6%, for setups on which only W, H or T remain operative. The performance decays to approximately 81% when A and K are the functioning sensors. The case in which all sensors are unavailable (e.g., a simultaneous battery discharge) is not presented here since it is obvious that no system could operate under such circumstances. In either case, this kind of situation is remotely seen. Finally, from the previous results it could be also concluded that some sensors are more important than others. In fact, sensors on the thigh and wrist appear to provide the most beneficial information, thus the highest impact of critical sensor anomalies is observed for configurations lacking of these sensors.

3.4.5. Tolerance to sensor faults

Conversely to what happens to occur to sensor failures, a faulty sensor is capable of delivering data. Nevertheless, sensor faults generally entail signal degradation. Some of the signal artifacts may be eliminated during the preprocessing stage, however, other anomalies may imply a certain information loss. The effect of this type of anomalies is here investigated.

As introduced in Section 3.2, when the sensor circuitry is not adequately supplied a reduction of the sensor dynamic range could be observed. This shortening translates into a change in the boundaries of the signal space, thus potentially leading to a misrepresentation of the actual body motion. For example, the sensors used in this study are capable of converting all measured accelerations within the range [-10g,10g]. Therefore, if the signal range is reduced to a tenth of the original interval (i.e., [-1g,1g]), most of the digitized signals will likely not represent the actual measured motion (see Figure 3.9).

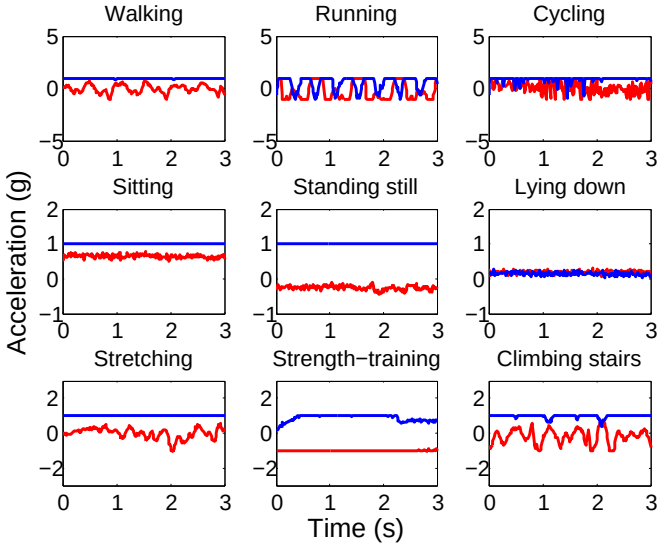


Figure 3.9: *Acceleration signals when the dynamic range reduces to a tenth of the original one (i.e., from ±10g to ±1g). Original acceleration signals are depicted in Figure 3.5.*

The tolerance of HWC and standard AR models to this phenomena is here analyzed. To that end, the systems performance is evaluated for two scenarios. In the first case, the dynamic range is reduced to a 30% of the original one (i.e., [-3g,3g]). In this new signal space active exercises are expected to be misrepresented, whereas low intensity activities will in principle not suffer relevant variations. In the second more challenging scenario, the dynamic range is reduced to a 10% of the predefined interval (i.e., [-1g,1g]). For this latter case, alterations in all activity patterns are devised. It must be noted that these scenarios are neatly selected after inspecting the considered dataset; highest acceleration values are slightly above 5g, therefore no relevant influence is expected when changing the dynamic range above ±5g.

Changes in the dynamic range may be simply modeled through a thresholding process (Figure 3.9). Concretely, those measures values that fall outside the bounds of the considered dynamic range are set to the interval extreme values. For the AR models, similar settings to the considered for the study of the tolerance to sensor failures are here

AR model\#faulty sensors	*0*	*1*	*2*	*3*	*4*	*5*
	New dynamic range = 30% original dynamic range					
SARC (hip)	81.52±4.56	66.15±3.78	-	-	-	-
SARC (wrist)	87.52±5.12	53.51±6.31	-	-	-	-
SARC (arm)	80.16±3.16	57.83±7.23	-	-	-	-
SARC (ankle)	82.52±3.79	58.16±8.12	-	-	-	-
SARC (thigh)	88.52±2.03	71.98±4.21	-	-	-	-
FFMARC	**97.39±1.69**	88.31±4.01	76.14±4.79	61.15±8.36	42.39±11.15	39.15±13.16
HD	89.84±2.57	85.35±4.16	79.77±8.96	68.21±13.15	59.17±16.14	52.75±20.07
MV	82.07±6.17	79.29±5.36	66.74±7.12	43.21±10.11	36.29±14.79	31.47±19.02
$HWC_{\alpha\beta\gamma\delta}$	96.34±2.34	**95.68±2.17**	**92.77±3.78**	**86.21±5.22**	**73.34±8.16**	**65.36±13.98**
	New dynamic range = 10% original dynamic range					
SARC (hip)	81.52±4.56	21.36±11.18	-	-	-	-
SARC (wrist)	87.52±5.12	17.78±9.37	-	-	-	-
SARC (arm)	80.16±3.16	26.31±14.13	-	-	-	-
SARC (ankle)	82.52±3.79	21.16±7.18	-	-	-	-
SARC (thigh)	88.52±2.03	19.98±6.41	-	-	-	-
FFMARC	**97.39±1.69**	69.16±5.39	41.26±8.12	17.23±15.13	21.07±10.86	18.19±8.94
HD	89.84±2.57	79.96±6.31	59.49±13.12	41.92±11.69	**29.75±17.25**	21.16±15.78
MV	82.07±6.17	77.16±6.01	46.19±11.16	38.21±9.98	27.18±12.87	**26.36±8.37**
$HWC_{\alpha\beta\gamma\delta}$	96.34±2.34	**94.23±1.79**	**86.77±6.03**	**53.21±21.84**	27.18±16.65	25.12±19.21

Table 3.2: *Average (standard deviation) accuracy values obtained for each ARC approach for diverse number of anomalous sensors. KNN and the ten features setting is used for standard and base classifiers.*

used (i.e., KNN and ten best ranked features). In fact, this configuration renders highest average performance for the diverse AR models in absence of anomalies (Figure 3.7). For the SARC model, the dynamic range of the corresponding sensor is modified. For MARC approaches, various configurations with an increasing number of anomalous sensors are evaluated. The faulty sensors are randomly selected from one iteration to another, but for the case in which all sensors are anomalous. The results obtained after evaluation are presented in Table 3.2.

For the case in which the dynamic range is reduced to a 30% of the original interval, a considerable performance worsening is seen for most AR systems. Models that rely on a single sensor are clearly the most sensitive to sensor faults. Nonetheless, practical differences are observed among the sensors considered in this study. Concretely, those sensors placed on body locations that are subject to lower accelerations demonstrate more robust, as it occurs for the hip and thigh whose performance decreases around 15%. On the contrary, wrist, arm and ankle sensors suffer from a higher reduction on their performance (35% at worst). This is motivated because sensors worn on the extremities are normally subject to higher accelerations, specially during the execution of intense activities such as running or cycling, and also walking to a lesser extent. These accelerations values are prone to fall out of the bounds defined for the new dynamic range. More tolerance to sensor

faults is gained when using multi-sensor configurations. Nevertheless, once again not all models behave similarly. FFMARC appears to be the most vulnerable MARC model to changes in the dynamic range, specially when two or more sensors malfunction. More than 20% drop with respect to the baseline is observed when two sensors are affected, 35% for three anomalous sensors and more than 50% for four or more faulty sensors.

DFMARC approaches turn to be the most consistent. MV demonstrates capable of dealing with changes on a sole sensor, but low tolerance to anomalies in a plurality of sensors (>40% performance drop from baseline). AR systems based on a HD approach appear to be more robust than those using MV. If top-ranked sensor classifiers are not affected by anomalies the performance is expected to be near to the baseline. However, as the number of erroneous sensors increases also grows the possibility of having a faulty high-ranked sensor. At worst conditions, the performance is observed to drop up to 40% from baseline. The high standard deviation values obtained for the HD approach could be explained since the anomalous sensor/s are randomly selected from one iteration to another. This leads to setups on which the designated faulty sensor/s may be either low-ranked (good HD performance) or high-ranked (poor HD performance). From all evaluated models, the HWC clearly outstands as the most robust approach to sensor faults. In fact, almost no worsening is detected when a minority of the sensors are affected, and only a 10% drop is seen when three faulty sensors are considered. The performance reduces to approximately 70% accuracy when the complete set of sensors functions defectively.

A much higher impact is seen when the dynamic range is shorten to a 10% of the default interval. Here, the performance of SARC models plummet to negligible values. FFMARC models also severely suffer from the effects of sensor faults. Already for one anomalous sensor the accuracy drops more than 25%, more than 50% for two affected sensors and nearly 80% when all sensors are faulty. The performance also declines more profoundly for the DFMARC approaches. HD shows a similar tendency to what was seen for the previous scenario. Then, if the sensor affected by the anomaly is low-ranked almost no influence is made on the final yielded decision, observing the opposite effect when the faulty sensor is high-rated. Anyway, very low performances are seen for two or more anomalous sensors. MV provides a reasonable tolerance to faults on one sensor, but no practical utility when two or more fail. Again, the most robust approach is the HWC, which shows almost no

worsening for an erroneous sensor, and a acceptable decline for two faulty sensors, thus confirming its potential use for dealing with sensor technological anomalies. Yet, under this complex scenario the model is not capable of overcoming the effects of a majority of faulty sensors.

3.4.6. Discussion

Performance in ideal conditions

This work does not aim at delving into the capabilities of AR systems devised for ideal technological conditions. In fact, to this respect much research has been performed during the last years and good models are available. The evaluation performed in Section 3.4.3 was rather planned to compare the recognition capabilities of the proposed HWC with respect to well-known standard AR approaches. Moreover, these results serves to this work as a baseline of the recognition capacity in absence of sensor anomalies. In the following the main characteristics of each model are succinctly described.

Systems based on a single sensor could provide an acceptable recognition performance, but at the expense of a rich feature vector. This translates into a more complex classification stage, thus potentially requiring more time for the system training and also increasing its latency during runtime use. Sensor fusion introduces a remarkable improvement with respect to the use of a single sensor. This is quite reasonable, since generally the more information from the body motion is available the better the user activity is normally described, albeit this might not be necessary for some specific actions (e.g., ankle sensor while detecting "hand-writing"). Concretely, FFMARC proves to be the most reliable approach from tested, specially when reduced feature vectors are used. Although a few features may be enough to provide an accurate solution, the size of the aggregated feature vector is proportional to the number of combined sensors, thus also increasing the complexity of the classification process.

Diverse interpretations could be extracted from the evaluation of the DFMARC models. In broad strokes, HD and specially MV demonstrate weak approaches when a reduced set of features is used. MV performance limitations come from the fact that all sensors have the same relevance into the decision process. When a majority of the sensors fail to recognize the activity, the eventually yielded activity is erroneous. In HD models the final decision normally relies on the high-ranked classification entities, therefore errors on top of the hierarchy are especially

critical. As individual sensor classifiers generally improve their recognition capabilities as the number of features increases, the overall performance of HD and MV models also grows. Through exploiting the main advantages of HD and MV models, i.e., a hierarchical and plural contribution of sensor decisions, the HWC manages to achieve a recognition characteristics similar to those obtained for the most reliable approach (i.e., FFMARC). For simple feature sets the HWC already proves to be quite accurate, which helps to reduce computational costs at the feature extraction stage. Even though, the HWC requires a considerable set of decision entities or base classifiers (N x M), which increases as the number of sensors (M) and activities (N) does. However, these are simple models that could further benefit from parallel computing, something that cannot be easily applied to other standard models. In either case, a fair comparison of the resources required for each model would be needed to this respect.

Tolerance to sensor failures

Major changes in on-body sensor setups are normally produced by critical sensor failures. At worst, sensors may get broken or damaged to an extent so they stop delivering data. Conceptually, similar situations may be observed when a user leaves a sensor behind, it gets out of battery or is powered down, resulting in a permanent loss of the signals. Under these circumstances, standard AR models devised for steady sensor configurations are prone to fail to provide activity-awareness capabilities.

From the analysis elaborated in Sections 3.2 and 3.4.4, single sensor approaches turn to be completely useless in the event of sensor failures. Since SARC models rely on the data collected through a particular sensor, if this fails to deliver data no recognition capabilities may be provided. Similarly, feature fusion multi-sensor approaches also become unusable in this context. FFMARC classification models are designed to operate on a particular input (i.e., aggregated feature vector), thus they cannot be naturally used if this input is incomplete. FFMARC models could be redesigned to operate with the remaining active sensors, however, this involves a complete retraining of the model, which is quite costly as well as impractical to be performed during the runtime use of the system. Therefore, AR applications based on SARC or FFMARC models present no other option than switching off the system and stopping the monitoring process until the setup is recovered to its

original or default form.

Halting the recognition process could be unacceptable for some applications (e.g., elderly fall detection, freezing of gait in Parkinson, or epileptic seizures detectors), and specially burdensome and discouraging for general users. Decision fusion models such as the one proposed in this dissertation are seen to be a valid solution to help not interrupt the AR process. In fact, since DFMARC models operate on the individual decisions provided by each sensor classifier or entity, modifications in the sensor network are in principle supported. Although applicable, HD and MV were qualitatively shown to be sensitive to these changes. Conversely, the HWC proves to be very robust to sensor failures. In fact, the model practically maintains baseline recognition capabilities even when a majority of the sensors are missing. At worst, when a sole sensor remains operational, the performance is similar or even higher to the obtained through a SARC approach, thus demonstrating the potential of the HWC also for single sensor setups, as well as its considerable scalability properties.

Not only should be the AR models capable of coping with the effects of occasional sensor failures but to facilitate user maintenance tasks. Thus for example, to provide a means to continue operating while a discharged sensor is being recharged is specially important in realistic applications. As concluded before, the HWC helps to provide seamless recognition capabilities, thus supporting quotidian real-world situations that may lead to a temporary absence of part of the utilized sensors.

Tolerance to sensor faults

Although less damaging than critical failures, sensor faults could also lead to a potential malfunctioning of AR systems. Conversely to the formers, faulty sensors are capable of delivering data, albeit this information is subject to degradation. This degradation normally corresponds to signal artifacts, some of which can be removed at the preprocessing stage (e.g., electronic noise could be filtered). However, other anomalies may entail a certain information loss that cannot be recovered through preprocessing techniques. This is the case of changes in the sensor dynamic range due to an inadequate energy supply of the device.

Variations of the dynamic range may suppose a misrepresentation of the measured body motion signals. Clearly, the most reduced the dynamic range becomes the higher the impact this anomaly has. Nevertheless, changes in the dynamic range produce a different impact on

each activity pattern. On the one hand, intense activities that involve a high body motion are primarily distorted, since their acceleration values may potentially fall outside the bounds of the anomalous sensor range. This is the case of running, cycling or walking for the activity set considered in this work. On the other hand, resting or low motion activities could remain unaffected if the variation of the dynamic range is not pronounced, but also distorted if this reduces sufficiently.

When sensors suffer from a moderate reduction of the dynamic range, the performance of SARC models considerably declines. The highest performance worsening is seen for those SARC models operating on data collected from body parts subject to intense accelerations (i.e., body extremities). This fits well with the explanation provided above to this respect. The use of various sensors may help overcome the effects of sensor faults, however, not all MARC models show similar robustness. FFMARC models are capable of partially coping with changes in one of the sensors, but show low tolerance to two or more faulty sensors. Artifacts introduced by individual faulty sensors contaminate the complete aggregated feature vector, therefore leading to misclassifications. The effects are more prominent as the number of anomalous sensors increases. DFMARC models benefit from the independence of each sensor classifier. In this way, HD and MV proves to be capable of facing the challenge of one faulty sensor, however, their recognition capabilities considerably drop when two or more sensors become anomalous. From all tested models, the HWC shows the best fault-tolerance. In fact, almost no worsening is observed when two or less sensors are affected, which keeps in reasonable levels also for three sensors. When a majority of the sensors are distorted, the performance notably reduces but it is still higher to what is achieved for other AR models in more beneficial circumstances.

When the dynamic range is more severely reduced, SARC and FF-MARC models show almost as useless as for the case of having sensor failures. HD and MV also show little resistance to the effects of the sensor anomalies, even when a single sensor is affected, thus of doubtful utility. Only the HWC offers an strong resilience to sensor faults, perfectly dealing with the situation of a faulty sensor and moderately coping with the effects of two anomalous sensors. Nonetheless, when a plurality of sensors are affected the HWC approach neither overcomes the effects of severe changes in the sensor dynamic range.

HWC advantages

The HWC was originally devised to cope with the effects of sensor failures and faults. To this respect, promising capabilities have been demonstrated along this discussion. Nevertheless, the HWC possesses other remarkable properties that are specially required in AR systems for the real-world. In the following these characteristics are described.

SARC and FFMARC models cannot function when a sensor breaks or disappears from the original sensor topology, whereas DFMARC models could leverage the information provided by the rest of active sensors. In this way, DFMARC models in general and the HWC in particular allow for an uninterrupted AR. Now, to return the system to its initial performance, the broken sensor must be replaced or substituted with a new one that may potentially have different characteristics (e.g., different calibration or signal modality). In this context, SARC and FFMARC could be newly utilized, but they require a complete retraining of the model. Conversely, decision fusion models only require to train the sensor classifier that operates on the new sensor. This training could be performed in a rapid fashion through expert models as will be investigated in Chapter 5. This is a very valuable characteristic in the AR domain, since depending on the problem complexity systems retraining may take a significant time.

The HWC also proves to scale well to the number of used sensors. As shown in Section 3.4.4, the HWC provides good results for diverse sensor setups, even for combinations of a reduced set of sensors or a sole single device. From this, not only does the HWC show to be useful for multi-sensor configurations but also applicable in single sensor setups. Anyway, these properties will be more deeply observed in Chapter 4, where more complex scenarios in terms of sensors and activities are evaluated.

The flexibility of the HWC not only applies at the sensor level but also at the activity level. AR systems are normally devised for a set of particular activities, however, this may change in the course of time depending on the particular user and application needs. For example, additional activities to the originally planned may be required when a new exercise routine is considered or a workout plan modified. These changes are not only seen to add new activities but also remove some of these at the point of need. This is found of special interest to reduce systems complexity and increase their recognition performance, as well as to procure systems personalization to subjects. Schemes based on a

standard AR model require a complete system rebuilding when the activity set is varied. Multi-decision or fusion techniques such as adaboost, decision stumps, random forests or other popular ensembles and meta-learners likewise require to utterly retrain the model. Conversely, the HWC supports this kind of reconfigurations. For the inclusion of new activities, only new base classifiers must be trained for the added activities, and their associated weights computed. If an activity is rather removed, an update of the model weights is only required. These properties are eligible to support important requirements of real-world AR systems such as self-configuration, auto-adaptation and evolvability. Section 6.3 shows next steps in this direction.

Two weighting models were proposed as part of this work, one corresponding to a unified weighting of each base classifier decision ($\text{HWC}_{\alpha\gamma}$) and a second model on which insertions and rejections are independently weighted ($\text{HWC}_{\alpha\beta\gamma\delta}$). Although both models show good classification properties, the second weighting approach demonstrates a higher potential. Through independently weighting insertions and rejections the HWC becomes more problem-sensitive, and capable of leveraging the use of all base classifiers even when their classification or rejection capabilities might be unbalanced. According to the weights, these could be defined through diverse criteria. In this work, accuracy ($\text{HWC}_{\alpha\gamma}$) and sensitivity-specificity ($\text{HWC}_{\alpha\beta\gamma\delta}$) metrics have been particularly considered, however, an important asset of this model is that it could be likewise operated by using other metrics or figures.

Open issues

In research, as important as it is to propose a new model or evaluate a technique is to compare these with previous works in the field. Unfortunately, the comparison with related work is here difficult since the effects of sensor failures and faults have been seldom investigated in the AR domain. Moreover, as mentioned in Section 1.3, there is no gold standard and also a clear lack of datasets for benchmarking AR models. To compensate all this, a comparison of the capabilities of the HWC with the most widely used AR solutions has been provided. Moreover, in order to ensure the reproducibility of our experiments, the models are evaluated on a dataset that has been extensively employed in this domain. Anyway, a strong effort must be put in the wearable AR domain to collect new datasets that may serve to validate these and other models.

The models used here to emulate the effects of sensor technological anomalies represent quite precisely what can be observed in a realistic setting. In fact, no differences are expected for critical anomalies or sensor failures. However, it would be of much interest to not only simulate their impact but further observe them in a real-world scenario. Unfortunately, this is not an easy task since sensor failures appear in a random and occasional manner. The closest approach to this respect could be the dataset collected in [105]. Here the authors gathered multimodal AR data on which sensors are sometimes switched off, but principally for energy saving reasons. Packets loss are also reported in this dataset, however, these are normally associated to missing data from turned-off sensors. In this direction, a long-term AR dataset including realistic sensor anomalies could be of much utility to further validate some of the results presented in this work.

The HWC model has been clearly demonstrated as the most robust approach. Nevertheless, for a plurality of faulty sensors the system reduces its performance, which may be more or less critical depending on the magnitude of the fault. To overcome this, a error detection procedure could facilitate to exclude the decisions yielded by faulty sensors. In this line, a recent work [55] proposed the use of distance measures and information theory techniques to identify erroneous measurements in a multi-sensor setup. The HWC could leverage this type of mechanism to not only identify the damaged sensors but also update the corresponding weights (γ_m and/or δ_m), thus lowering their impact on the eventual yielded decision.

Not only could changes in the sensor setup be incorporated in the HWC model but also at the activity level. Sensor anomalies may affect the recognition of part of the activities (e.g., intense activities when the dynamic range is reduced), but not alter the identification capabilities for others. Then, instead of reducing the decision weight at the network level (γ_m, δ_m) this could be rather performed at the sensor level (α_{mn}, β_{mn}), thus only the weights of those base classifiers associated to unrecognizable activities are modified. This updating procedure is not only devised to overcome the limitations imposed by sensor technological anomalies but may be also utilized to dynamically adapt the AR system to people changing conditions. This is discussed in detail in Section 6.3.

3.5. Conclusions

Classic AR systems assume that the sensor setup remains invariant during the lifelong use of the system. Nevertheless, as other electronic devices, on-body sensors are subject to faults and failures. These technological anomalies lead to changes in the sensor data streams, which are normally unforeseen during the design phase and unpredictable at runtime. Consequently, models trained on ideal signal patterns may react in an undesired manner to anomalous sensor data. This potentially translates into a partial or total malfunctioning of the AR systems.

Sensor technological anomalies are categorized in sensor faults or failures depending on their criticallity. Faulty sensors are capable of delivering data, although normally containing artifacts that translate into signal degradation. Sensor faults could be produced by decalibration, parasitic circuit effects or an inadequate energy supply. More deleterious are sensor failures since they imply a total loss of information. Sensor failures are normally motivated by critical sensor breakdowns or damages, or simply appear when a sensor gets out of battery or shuts down. Similar effects could be seen when a user leaves a sensor behind or forgets to switch it on because for all these cases no data is available. In this chapter, the impact of sensor technological anomalies on standard AR solutions has been investigated, and a new method to deal with the effects of these anomalies has been proposed.

From the analysis of failure tolerance of standard AR approaches, it can be concluded that these are in general incapable of coping with the effects of sensor failures. AR systems based on a single sensor (SARC) cannot operate when the device stops delivering data, for example, simply when the sensor gets out of battery. The use of multi-sensor setups may help to overcome this limitation through leveraging the remaining active sensors. However, in practice not all sensor fusion models are capable of dealing with failures. In fact, feature fusion models (FFMARC) show no tolerance to sensor failures. FFMARC aggregates all features extracted from each sensor in a single vector, which is input to a given classifier. This classifier is trained at design time to work on this predefined feature structure. Thereby, if some sensor data are missing, the feature vector is incomplete and the model incapable of providing recognition. Conversely, decision fusion models (DFMARC) combine the individual decisions yielded by each sensor classifier, thus supporting a continuity of recognition through the use of the rest of unaffected sensors. Nevertheless, it is shown that HD and MV, the two

possibly most used DFMARC models, present special limitations that make them specially sensitive to sensor failures. HD performance deteriorates when has to rely on average sensor classifiers while MV suffers from the "tyranny" of a plurality of weak sensor classifiers.

Some sensor faults could be solved during the preprocessing stage (e.g., spurious spikes or electronic noise removal through a filtering process). However, other anomalies imply information loss that cannot be solved through signal processing techniques. This is the case of variations in the dynamic range, normally associated to the malfunctioning of the batteries or the irregular supply of the sensor circuitry. Upon assessing the robustness to dynamic range faults, standard AR systems demonstrate little capacity to deal with their effects. The recognition capabilities of SARC approaches severely degrade, specially for those relying on sensors placed on body parts subject to high motion. Likewise, FFMARC models show an important performance worsening since the anomalies introduced by a sensor contaminate the complete feature vector built during the aggregation process. The performance degradation is seen to increase as the number of anomalous sensors grows. Finally, once again DFMARC models present the highest tolerance to sensor anomalies, however, this reduces with the increase of the amount of faulty sensors.

From the observation of the lacks of classic AR approaches, an original alternate model to deal with sensor technological anomalies has been presented. The HWC leverages the benefits of HD and MV models and avoids their limitations. For each particular sensor, base classifiers are defined for the identification of each activity class. The decisions provided by these classifiers are weighted according to their recognition capabilities. The weighted decisions are fused for each sensor. The decisions adopted at sensor level are further weighted and combined to yield the eventual recognized activity. In ideal conditions, the HWC renders a performance quite similar to the FFMARC, and far higher than SARC, HD and MV models. Even more important is that the HWC proves to deal with sensor failures under which conditions SARC and FFMARC do not work at all, and HD and MV show low robustness. Moreover, the HWC demonstrates a high fault tolerance when a minority of the sensors are affected. Nonetheless, when a plurality of sensors are affected the HWC approach does not overcome either the effects of severe sensor faults. Detecting sensor faults could be of much utility to temporarily leave these faulty sensors out of the inference process. The flexibility of the HWC model may support these situations

through the dynamic reconfiguration of the sensor setup.

Other properties that facilitate the seamless use of the recognition system have been identified for the HWC. For example, the model may be easily modified to recognize new activities or adapt to user changing conditions. Likewise, the HWC flexible structure can support the sensor network reconfiguration at runtime. Through this, hot swaps or setup changes are in principle allowed, only requiring a training of the newcomer device and an update of the model parameters to the new sensor setup. A novel method to support the online training of newcomer sensors is presented in Chapter 5.

4

Robustness of AR systems to sensor deployment variations

4.1. Introduction

Technologies of daily living are devised to facilitate users normal activity and safeguard their welfare. In the last decades, thousands of products have been released to make our life easier, safer and more comfortable, thus allowing us to concentrate on more important tasks. In this context, wearable technology appears to provide new services to empower people in their habitual tasks in a transparent manner. To that end, sensors and systems are made part of articles of everyday use, principally embedded in accessories. However, transparency is not only achieved through concealing technology in a physical manner but when no influence on the human normal behavior is attained. In fact, on-body systems cease to be transparent when users start needing to wear the sensors in a particular manner, for example, when a bracelet must be worn on a specific limb or a watch positioned in a determined orientation.

To implement activity recognition, current systems normally require that sensors must be attached at predefined positions to discriminate between different actions. Pattern models are derived in a training step before the system deployment, where the sensor positions are considered to be constant. In particular for on-body sensors, a constant position on the body cannot be maintained in real-life scenarios. Sensor deployments are subject to variations introduced by the people normal use of the accessories into which these sensors are embedded. These variations correspond to sensor position changes or displacements at the user's body. Displacements can remain static during the execution of many activity instances, e.g., when sensors are misplaced and kept that way for the whole day. Sensors may further be dynamically displaced due to the effect of loose-fitting attachments, e.g., when integrated into baggy clothes. The effect of placement-related changes on sensor measurements is profound. Compared to expected patterns, sensor data distributions can change widely and along extended time spans. As a consequence, previously trained pattern models may fail to identify actions in the observed sensor data.

4.2. Sensor displacement

Sensor displacement stems from the position variations that a sensor experiences when moved with respect to an initial given placement. In on-body sensing, sensor displacement is observed when a sensor mounted

on a given body part is moved to another body location, which may be close (small to moderate displacement) or distant (extreme depositioning) to the initial placement. Displacement related sensor anomalies are seen to be highly relevant for inertial sensors, thus in this work profoundly analyzed.

Sensor displacement may be seen as the combination of two transformations: rotations and translations. According to the physics of the rigid body[1], rotations refer to the circular movements that the sensor experiences around its rotation axes or upon itself. Translations correspond to the movements of the sensor from a given position to another distant position through a specific direction. While sensor displacement between different limbs are less common in real-life applications, shifts and rotations on the same limb occur frequently.

Sensor displacement applies to each inertial sensing modality (acceleration, rate of turn, magnetic field) to a different extent. Thus for example, acceleration is specially sensitive to rotations. Rotations introduce a change in the sensor local frame of reference with respect to its original spatial distribution. This causes a shift in the direction of the gravitational component with respect to the sensor reference frame. The effect of translations is normally more dependent on the initial and end position as well as the magnitude of the acceleration experienced by the sensor. Thus for example, a sensor which is displaced from the upper arm to the lower arm will generally measure higher acceleration. On the contrary, during inactivity or while idling this change may not have appreciable consequences. More robust to displacement anomalies are gyroscopes, which are minimally affected by rotations along their rotation axis and translations along the same body limb. Gyroscopes do not measure exact angles but angular velocity which is integrated to obtain angular positions. In case of motion, the gyroscope remains unaffected by displacements along a given body limb since during a translation or rotation all points of the rigid body are rotated or moved to a similar extent (same linear and angular velocity/acceleration for all points of the same body limb). However, gyroscopes provide no relevant information when the user remains still, thus proving of little utility to assess sedentary or passive activities. The compass (magnetic

[1] A common approximation used in biomechanics or human body motion modeling consists of rigid segments (mostly representing body limbs) which are connected through joints that allow rotation around one (e.g., elbow) or more (e.g., wrist) axes. Although the human body cannot be literally seen as a rigid body (soft tissue, skin motion, muscle activity) this model fairly approximates most motion interactions and provides a tractable approach for analysis and representation.

field sensor) measurements are also affected by rotations and to a lower extent by translations when assuming no gimbal lock degeneration[2].

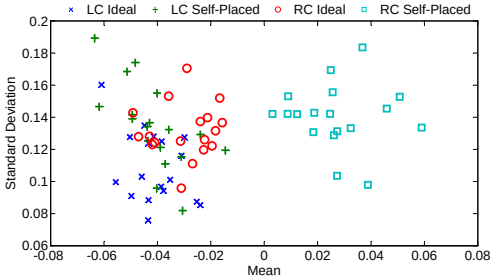
Sensor displacement leads to a new sensor position which results in a change in the signal space. The impact of the displacement on the sensor signal may vary depending on several factors, such as the magnitude of the displacement or the body part considered. Likewise, these displacement effects are also subject to the particular activities, gestures or movements the user performs. For example, a higher acceleration may be measured during running exercises when a sensor is displaced from the upper arm to the wrist, however, for this very case smaller accelerations could be registered when the user performs strength-training exercises such as push-ups. In either case, the sensor readings in the new signal space likely differ with respect to those expected from a default or predefined sensor placement. These changes propagate through the different stages of the ARC, thus affecting the inference process. An example of such effects is depicted in Figure 4.1. Here, a sensor displacement is unintentionally introduced by the user when self-attaching the devices (Figure 4.1(a)). This displacement translates into a significant drift at the feature level (Figure 4.1(b)). This shift in the feature space complicates the posterior reasoning process. Therefore, a model trained under the assumption of a predefined placement of the sensors (and accordingly a bounded feature space) may not correctly operate due to the variations introduced by the new feature space.

4.3. Synthesizing sensor displacement

To investigate the effects of sensor displacement requires actual situations on which these displacements manifest. Unfortunately, most of the datasets used in AR do not account displacement issues, i.e., all assume fixed sensor deployments [10, 106, 107, 108, 109, 110]. Then, these data are in principle useless for the sake of this study. However, an alternative is proposed in previous work, which consists in modeling the effects of sensor anomalies by introducing synthetic variations into the original recorded data. These variations try to emulate the trans-

[2]Gimbal lock refers to the loss of one degree of freedom within a gimbal system - pivoted support that allows the rotation of an object about a single axis - in a three-dimensional space that occurs when two out of three gimbals line up in a parallel configuration, “locking” the system into a rotation in a degenerated two-dimensional space.

(a) Sensor displacement originated during a user self-placement of the sensors



(b) Associated feature drift

Figure 4.1: *Example of sensor displacement introduced during the user self-placement of a sensor (**a**) and its effect at the feature level (**b**). In this particular example the displacement from the predefined deployment to the self-placement case applies to the right calf (RC) while the placement remains approximately similar for the sensor attached to the left calf (LC). In (**b**) the mean and standard deviation computed from the sensor acceleration signals is represented for various instances of a given activity.*

formations experienced by the default signal space as consequence of the sensor positioning change.

By following the aforementioned approach, this section aims at analyzing the effects of sensor displacement on standard AR systems, particularly comparing their impact on systems based on a single on-body sensor and for those of multiple body parts sensing. Likewise, the HWC model proposed in Chapter 3 is here also evaluated as a possible means to deal with the effects of sensor displacement. Additionally, this section aims to gain insight into which body locations are most robust

to the effects of sensor displacement as well as how each component of displacement, i.e., rotation and translation, contributes to these effects.

4.3.1. Rotational and additive noise models

The modeling of sensor displacement requires to emulate both rotations and translations. From the literature, these could be approached through the so-called rotational (RN) and additive (AN) noise [51, 53]. Here the noise does not refer to the classical idea of "unwanted data without apparent meaning" but to an unexpected by-product of the sensor positioning changes when a fixed setup is presumed. Moreover, RN and AN models are hereafter defined only for acceleration data. Other synthetic models would be necessary for different signal modalities. At any rate, RN and AN must be seen as approximations to the actual sensor displacement phenomena.

According to the physics of the rigid body, sensors may experience a rotation in the 3-dimensional Cartesian space. The rotation implies a change in the sensor local frame of reference with respect to its original spatial distribution. To refer the original signals to this new coordinates system an Euclidean transformation is used. This transformation is modeled through a rotation matrix[3] defined by the Euler angles ($\phi = \phi_{RN}, \theta = \theta_{RN}, \psi = \psi_{RN}$) which represent the rotation along the axes of the original frame of reference:

$$M_{RN} = \begin{bmatrix} c(\theta)c(\psi) & -c(\phi)s(\psi) + s(\phi)s(\theta)c(\psi) & s(\phi)s(\psi) + c(\phi)s(\theta)c(\psi) \\ c(\theta)s(\psi) & c(\phi)c(\psi) + s(\phi)s(\theta)s(\psi) & -s(\phi)c(\psi) + c(\phi)s(\theta)s(\psi) \\ -s(\theta) & s(\phi)c(\theta) & c(\phi)c(\theta) \end{bmatrix} \tag{4.1}$$

The translation of the original signals $(x_{raw}, y_{raw}, z_{raw})$ into the rotated coordinates system $(x_{rot}, y_{rot}, z_{rot})$ is obtained through:

$$\begin{bmatrix} x_{rot} \\ y_{rot} \\ z_{rot} \end{bmatrix} = M_{RN} \begin{bmatrix} x_{raw} \\ y_{raw} \\ z_{raw} \end{bmatrix} \tag{4.2}$$

The modeling of a sensor translation is not as "trivial" as for the rotation case. Sensor translation not only depends on the magnitude of the translation but also on the original location and its direction. For example, when sensors originally devised for the upper part of the extremities (shoulder/thigh) replace on the lower part (wrist/ankle),

[3] $c()$ and $s()$ represent cosine and sine functions respectively.

higher acceleration could be measured. An attenuation might be conversely observed when the sensors are displaced the other way around. Nevertheless, this behavior pattern is not always generalizable since it varies with the particular executed actions as commented in Section 4.2. The problem could be even more challenging if the sensors are depositioned from their predefined placement to a completely uncorrelated body part (e.g., from the ankle to the wrist).

Bearing in mind the complexity of the modeling of sensor translation it is here approximated through a stochastic model. Concretely, a similar model to the proposed in [53] is considered, which is based on additive white Gaussian noise characterized by a zero mean and with the strength of the anomaly in the value of the variance (σ^2_{AN}). Even if this model is not as precise as the defined for the rotational noise, the additive noise may model the casuistry of the problem as well as emulate some of the expected changes in the signal space. Besides, this kind of model could be particularly suited for the modeling of multiple translations along a given direction originated from a loose-fitting sensor.

4.3.2. Experimental setup

To analyze the robustness of AR systems to sensor displacement, similar systems to the ones studied in Chapter 3 are here evaluated[4]. Thus, the very setup and dataset used in previous experiments are here also employed (see Section 3.4.2). In short, the used dataset comprises acceleration data registered for 20 subjects while performing a set of daily living activities. From the complete set of activities the most representative nine are selected. AR systems based on a single or individual sensor and those of a multi-sensor configuration are respectively tested. For the single sensor approach a sole ARC is required (i.e., SARC). For the multi-sensor system, feature fusion (FFMARC) and decision fusion (DFMARC) approaches are followed. For the decision fusion approach, the HWC structure is utilized, concretely the $\text{HWC}_{\alpha\beta\gamma\delta}$ given its outstanding capabilities. From the evaluation performed in Section 3.4.3, the configuration that yielded the best performance in ideal conditions for the different AR models is here employed. Thus, the ten best features ranked from an original set of up to 861 features [101] are consid-

[4]All the processing is performed in Matlab R2011b. For the preprocessing, featuring and classification stages some of the functions provided in the *Signal Processing*, *Statistics* and *Bioinformatics* toolboxes have been used, while many others have been specifically defined for this purpose.

ered. For that search a feature selector based on the receiver operating characteristic is used [95]. KNN models similar to the ones used in prior work of this thesis are used as classification paradigm [103]. The coefficient k is empirically obtained and set to three since this rendered the best performance. A ten-fold random-partitioning cross-validation is performed to support an adequate comparison among the models (Section 2.1.7). The experiment is repeated 100 times to ensure statistical robustness.

Concerning the use of the displacement models presented in Section 4.3.1, it should be first considered the particularities of the sensors used in this evaluation. Concretely, biaxial (XY) accelerometers are used to register the body motion, thus only two rotated signals may be obtained from Equation 4.2. Thus, given x_{raw} and y_{raw}, and considering $z_{raw} = \bar{0}$, the signals after introducing rotational noise are:

$$\begin{aligned} x_{rot} &= [c(\theta_{RN})c(\psi_{RN})]\, x_{raw} - [c(\phi_{RN})s(\psi_{RN}) + s(\phi_{RN})s(\theta_{RN})c(\psi_{RN})]\, y_{raw} \\ y_{rot} &= [c(\theta_{RN})s(\psi_{RN})]\, x_{raw} + [c(\phi_{RN})c(\psi_{RN}) + s(\phi_{RN})s(\theta_{RN})s(\psi_{RN})]\, y_{raw} \end{aligned} \tag{4.3}$$

On the contrary, the additive noise is applied independently of the number of dimensions (axes) the sensor has. An example of the effects of the modeled sensor displacement could be seen in Figure 4.2.

The particular considered procedure consists in the application of each respective noise model to each individual instance or data window. Only test instances are subject to RN or AN since the training data are assumed not to comprise such variations (i.e., predefined AR system). The rotation angles $(\phi_{RN}, \theta_{RN}, \psi_{RN})$ are arbitrarily varied for each instance from 0 to a maximum value $\angle_{RN}$ for the rotational noise, while the noise variance coincides with the level defined by σ^2_{AN} for the additive noise.

4.3.3. Single sensor performance

The performance for the single sensor-based recognition models under the effect of rotational and additive noise is depicted in Figure 4.3. Baseline recognition performance is shown for $\angle_{RN} = 0°$ and $\sigma = 0$g, which correspond to ideal circumstances. For both rotations and translations, it is observed that the higher the noise (displacement) introduced the lower the accuracy of the recognition system is.

Particularly, for the case of the rotational noise (Figure 4.3(a)) slight sensor variations ($\angle_{RN} \leq 15°$) are normally well tolerated. However, a significant drop is observed for sensor rotations of 30° or more. Clearly,

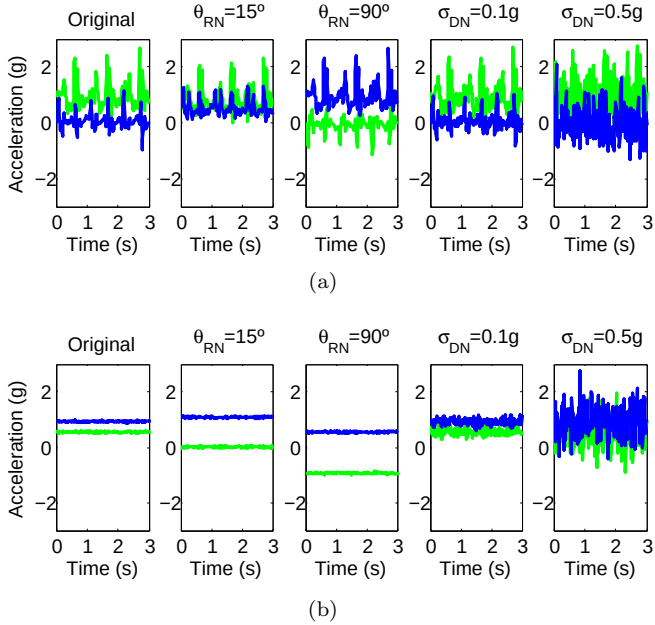


Figure 4.2: *Effect of the rotational and additive noise. X-axis (green) and Y-axis (blue) accelerations recorded through the hip sensor when (**a**) walking and (**b**) sitting. Legend: 'Original' ≡ raw signals, '$\phi_{RN} = \theta_{RN} = \psi_{RN}$' ≡ data with rotational noise (in °), 'σ_{AN}' ≡ data with additive noise (in g).*

as may be concluded from Equation 4.3, the effect of the rotation is more profound as higher its value. For example, for θ_{RN}=90° x_{rot} turns to be a scaled version of y_{raw}, with ϕ_{RN} and ψ_{RN} determining the scale factor. Therefore, the signal measured on the X axis has almost no relation with the signal that would be measured in absence of displacement. As for this example, there are infinite combinations of ϕ_{RN}, θ_{RN} and ψ_{RN} leading to infinite transformations of the original signal space. This translates into variations of the feature space that cannot be followed by the predefined SARC model.

A particularized analysis of the performance for each sensor drives us to think that hip and thigh sensors are in principle the less robust to rotations. This is probably because the magnitude of the acceleration registered in these positions is lower than for the rest of sensors, thus

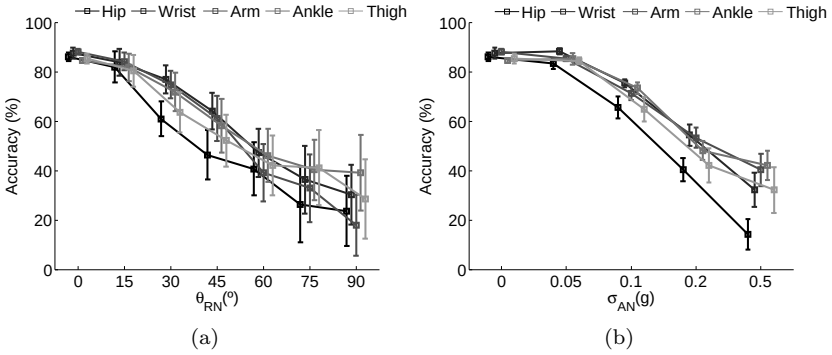


Figure 4.3: *Effect of the (**a**) rotational and (**b**) additive noise on the performance of predefined SARC models. Each SARC model operates on the data registered through each individual sensor. The error bars along the curves correspond to the standard deviation of the recognition accuracy.*

defining more sensitive classification boundaries. On the contrary, wrist and ankle sensors happen to be the most robust to rotations.

The performance of the systems remains practically unaffected when a negligible level of additive noise is applied. This might correspond to a subtle translation. However, when the noise level is increased up to $100mg$ or more the systems are incapable of accurate recognition. Quantitatively, this translates into a 15-20% average performance drop with respect to the accuracy obtained in normal circumstances. The drop further grows above 40% and 60% when σ_{AN} is increased up to $200mg$ and $500mg$ respectively.

As for the rotational noise, the most sensitive sensors are those placed on hip and thigh. A similar explanation to the one provided for the rotational noise may be likewise used here. Therefore, it can be in principle stated that hip and thigh locations are less recommendable (in terms of robustness to displacement) for AR solutions based on a single sensor.

Finally, it is also important to stress on the increasing standard deviation values obtained as the level of noise grows. Different displacement settings are tested since the rotation is varied between 0 and $\angle_{RN}$ for each experiment (iteration), thus driving to diverse performance results. This provides performance bounds (best and worst) for

the impact of sensor displacement. For the case of the additive noise the standard deviation proceeds from the stochastic model considered.

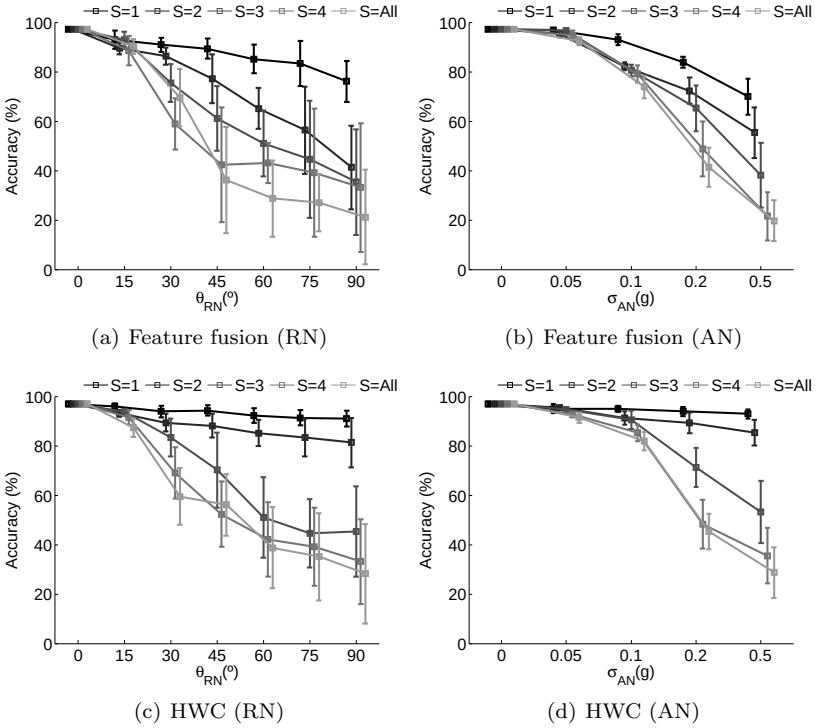
4.3.4. Multi-sensor fusion performance

A similar study is presented in Figure 4.4 for the multi-sensor approach. Here both feature fusion and HWC approaches are tested to evaluate the tolerance of these models to displacement. Now, a subset S of the complete set of sensors is affected by the corresponding noise for each case. The displaced sensors are randomly selected from one iteration to the next in order to guarantee that different combinations of sensors are tested, thus reflecting what may happen in a realistic scenario.

Significant differences may be seen with respect to AR systems of one individual sensor. As already demonstrated in Chapter 3, multiple sensor configurations perform normally better than single sensor-based models. Apart from that, fusion approaches also tend to be more robust to sensor anomalies. However, feature fusion and decision fusion models behave differently in presence of sensor rotations.

The feature fusion model (FFMARC) partially copes with the effects of rotations of a sole sensor. At worst, the performance drops around 20% with respect to ideal circumstances. Also rotations of two sensors may be assumed but for moderate figures ($\theta_{RN} \leq 30°$), in which the performance is found to worsen 15% or less. However, the worsen is more dramatic when rotations of higher magnitude are considered and/or more sensors are rotated. Thus for example, for S=2 and $\theta_{RN} \leq 45°$ the average recognition performance is below 80%, which progressively declines to almost 40% for rotations of 90°. The results are even much worse when three or more sensors are displaced, leading to recognition systems of limited utility when sensor rotations of 30° or more are applied.

The performance of the HWC model remains practically the same independently of the level of noise added when one sensor is rotated. This is due to the fact that the decisions yielded by the rest of the sensors cope with the failures introduced by the disturbed one, thereby allowing for a performance almost similar to what is achieved in normal circumstances. The model is also able to satisfactorily overcome the challenge of two rotated sensors, with a performance drop inferior to 10% for θ_{RN}=45°, and 20% at worst conditions (i.e., θ_{RN}=90°). However, as the number of rotated sensors increases the probability of misclassification grows. This leads to certain cases where the fusion



(a) Feature fusion (RN)

(b) Feature fusion (AN)

(c) HWC (RN)

(d) HWC (AN)

Figure 4.4: *Effect of the* ***(a,c)*** *rotational and* ***(b,d)*** *additive noise on the performance of predefined sensor fusion models.* ***(a,b)*** *correspond to a FFMARC approach while in* ***(c,d)*** *the HWC is used. S identifies the number of sensors simultaneously 'displaced' through the respective noise.*

performance severely worsens. In fact, this may be seen when a majority of the sensors are rotated ($S \geq 3$), with a decreasing performance which nevertheless overtakes the achieved for SARC and FFMARC approaches.

The above analysis could be also extended for the case of the emulated translations. Low levels of additive noise do not affect much the performance of fusion approaches. FFMARC tolerates well moderate noise levels (i.e., σ_{AN}=0.1) when a sole sensor is affected. However, the worsen is much more pronounced for the case of two or more anoma-

lous sensors. For this case, the performance may drop from 60% to 80% for most noisy conditions and a majority of the sensors affected. The HWC model shows a higher robustness, even for considerable noise levels. Actually for $S \leq 2$ the system performs almost perfectly, but a significant level of noise ($\sigma_{AN} \geq 0.2$) on three or more sensors results in an important performance worsening. In any case, the HWC model outperforms once again SARC and FFMARC models, thus demonstrating the usefulness of this approach.

4.3.5. Discussion

Single vs. multiple sensing

AR systems based on a single sensor have been demonstrated worthless but for very small displacements. Moreover, the impact of sensor displacements has been shown to depend on where the sensor is originally attached as well as the particular activity performed. Those sensors located on body parts subject to a low motion (e.g., hip, thigh) normally measure lower levels of acceleration, thus being specially indicated for the discrimination among sedentary activities. In such circumstances the static acceleration basically dominates, thus the effect of the displacement becomes particularly relevant, especially for the rotations. For example, a sensor placed on the hip may help to discriminate between activities such as standing or lying down, but if the sensor is significantly rotated the system may interpret the former activity as the latter and vice versa. On the contrary, sensors located on the extremities (primarily the upper ones) appear to suffer from sensor position variations to a lower extent. In this regard, the wrist sensor turns to be the most robust of the whole set of analyzed on-body placements. Differences on the translation effects among different body parts are not that clear to interpret. At first glance, the effects of translations are expected to be more critical when the sensor is placed on a limb. Nevertheless, this may vary depending on the actions the user carries out. This means that the results are in principle activity-dependent.

The use of multi-sensor configurations demonstrates highly recommendable to counteract the effect of sensor displacement anomalies. FFMARC models cope with the limitations introduced by low to moderate rotations and translations, but only under the assumption of one rotated sensor. Nevertheless, as the number of affected sensors increases the effect of rotations and translations turn to be more harmful than for single sensor configurations.

The HWC model proves to be the most robust approach to sensor displacements. For a minority of displaced sensors, and independently of the magnitude of the displacement, the HWC model maintains reasonable recognition capabilities. For some cases, the performance is almost similar to what is achieved in ideal circumstances. In fact, practically no drop in the performance is observed when just one sensor is displaced. The tolerance of the HWC to sensor displacement is seen to depend on the magnitude of the displacement when a majority of the sensors are displaced. Thus a significant worsening is observed for large rotations and translations.

HWC advantages

The use of multiple sensor configurations provides higher tolerance to sensor displacement. Fusion approaches are meant to combine the data delivered by these sensors, however, not all aggregation models are demonstrated to be robust to displacement. Feature fusion defines a single feature vector with the data coming from each sensor. Any variation in the signals (even for only one sensor) is further propagated to this vector, thus introducing a change into the original feature space. As for single sensor-based AR systems, this reduces the detection accuracy. HWC overcomes this limitation by treating each sensor data individually. Therefore, the fusion of the decisions yielded by each sensor instead of the aggregation of their features allow us to cope with variations in a minority of the sensors.

A quantitative comparison with other related approaches is difficult since different setups, datasets and methodologies are considered for each case. In addition to that, there are very few contributions that analyze the effects of sensor displacement as indicated in Section 2.4. From these, [51] is here considered for comparison given the similarity to our approach. In [51] the authors showed that simple decision fusion models such as MV may be used to deal with rotation and faults anomalies when a significant number of nodes (up to 19) is considered. Nevertheless, the performance significantly decreases for reduced sensor networks. As demonstrated in previous work of this thesis, MV suffer from serious limitations specially for reduced sensor networks. In this regard the HWC model offer a more scalable solution with remarkable results even when just five sensors define the whole network. Moreover, the study performed in [51] is limited to the analysis of upper body extremities, and no analysis is provided to identify which body parts

are more sensitive to sensor displacement or other anomalies. Finally, it is also important to highlight that the results shown in [51] correspond to a subject-specific (in fact just one) and instance-based setting, with multiple subsequent repetitions of a specific gesture. These realizations define a much more compact signal space where the effect of displacement is more evident. Our approach has been conversely tested on a more general context, for a higher number of subjects and for a more realistic daily living scenario, yet displacement anomalies have been in both studies synthetically introduced.

Design and evaluation tool

Figures such as the ones obtained from the evaluation of displacement effects (Figure 4.3-4.4) could be "reversely" used to identify the number of affected sensors in a given deployment as well as the magnitude of the displacements. On the one hand, given the performance of the AR system, $\angle_{RN}$ or σ_{AN} could be identified. On the other hand, once fixed the level of RN and AN, the amount of displaced sensors could be extracted for multi-sensor configurations. More importantly, this might also serve to identify which sensor (i.e., body part) is subject to displacement if a single sensor setup is used. Anyway, it must be born in mind that the use of these results for designing purposes is particularly devised for AR problems that consider activities of similar nature to the ones here evaluated.

Similar figures could be also used during the recognition system design phase for those cases in which the level of displacement could be foreseen and is somehow constrained. In either case, further evaluation would be required to provide more precise problem-specific figures for this purpose.

Displacement effects

From the two sensor displacement components, rotations seem to produce the higher worsening of the AR systems performance. Small rotations have been shown to already have a severe impact on the systems recognition capabilities, specially for single sensor approaches. According to the synthetic model for rotations, a drift of the sensor frame with respect to its predefined orientation may lead to offset variations, signal attenuation or amplification, or even more harmful changes in the signal pattern. Small sensor translations introduce little changes with respect to the default signal patterns, which can be faced by most recognition

systems. Even though, the models demonstrate quite sensitive to large translations (i.e., high additive noise levels).

During the experimental evaluation of sensor displacement, a fixed displacement was not applied but rather dynamically varied. The rotation angles were arbitrarily changed for each instance from 0 to a maximum value $\angle_{RN}$. The additive noise also varied for each iteration owing to its stochastic character. Therefore, the results presented along this section may not only describe the effects of static displacement but also dynamic sensor position variations, such as the observed when the sensors are embedded in loose-fitting clothes or accessories. In either case, synthesizing displacement effects has important shortcomings that are next analyzed.

Open issues

Emulating sensor displacement in a synthetic manner may serve us to provide a first idea about how sensor displacement affects the normal behavior of AR systems. Moreover, synthetic approaches allow us to evaluate multiple scenarios and conditions, which are difficult to capture in a real setting. However, this approach lacks of various aspects. The rotational model applies well to accelerations, specially for low intensity movements. Nevertheless, the fact that both static and dynamic components are superimposed determines this is an average transformation. The additive noise, although also used in previous works, does not completely describe sensor translations since it may also represent other possible anomalies. Moreover, the models applied to the acceleration data are in principle not suitable for other sensing modalities. The modeling of rotations and translations for other inertial sensors such as gyroscopes or magnetometers is considered to be more complex than for the acceleration model. Thus, RN and AN should be redefined for the evaluation of displacement effects on these other modalities.

The dataset used here comprises 2D acceleration data, thus the synthetic models could not perfectly represent actual sensor displacements. The rotated signals are summed scaled versions of the signals registered under absence of displacement. Thus, when one of the components (here the Z-axis) is missing the estimated signals may be somehow biased. Even when from a mathematical perspective this is completely correct, practical differences could be seen when compared to a real setting. Conversely, the additive noise model works well with independence of the number of axes the sensors have since this transformation is indi-

vidually applied to each signal.

Sensor displacement has been here decomposed into two different problems, rotations and translations. Although this is interesting to analyze the impact of each displacement component and applies well when only one of them affects the sensor, in daily living settings a combination of these may be normally expected. Therefore, a model that includes both rotations and translations is particularly required for an extended evaluation of sensor displacement effects.

For the multi-sensor approach it should be also considered the unlikely event in which most sensors are displaced, case where the problem of the performance worsening is not completely solved. For that reason other additional mechanisms could be possibly considered. One approach is increasing the robustness of the recognition system through the definition of more robust data distributions. The idea would be to emulate the possible changes in the signal space and include them in order to be learned during the system training. For example, for the problem associated to sensor rotations one may think about training the system with artificially modified data for different magnitude rotations. Another alternative would be to collect data for a sufficient number of representative possible rotations, but it results almost impossible to fully cover all the potential rotations that a sensor may suffer from its original position. An investigation of real-world observations of sensor displacements may nevertheless reveal more about the impact of sensor displacements. In the following, an extensive investigation is performed in this direction.

4.4. Realistic sensor displacement

Section 4.3 presents a complete evaluation of the effects of synthetically modeled sensor displacements on diverse types of AR systems. However, synthetic models are an approximation to the real phenomena, thus they may not perfectly represent the actual effects of sensor displacements. Sensor rotations could be moderately well emulated, however, translations have demonstrated to be more complex to model. In fact, the additive noise model used to approximate sensor translations may not only represent the effects of translation but also other sensor anomalies. Therefore, it is seen necessary to further extend the analysis performed for the synthetic case to a more practical setting.

The effects of sensor displacement in real-world conditions are analyzed in this section. To that end, a complete benchmark dataset ac-

counting for realistic sensor displacements is collected, to date missing for the study of this problem. Various concepts of displacement are investigated, including those originated from the user self-positioning of the sensors or from a large sensor depositioning. Moreover, a means of introducing sensor displacement anomalies in a practical and structured fashion is also presented. The effects of displacements are analyzed from both statistical and recognition performance points of view. Systems akin to the ones used in previous sections are here considered for evaluation.

4.4.1. Implementing realistic sensor displacement

Although there exist multiple datasets to benchmark AR systems, there is no single dataset that widely investigates the effects of sensor displacement. To the best of our knowledge only a few works approached the problem through "realistic" data, but for a specific body limb and for in-lab conditions. These few datasets are further constrained in terms of number of subjects, activities and even the type of displacement considered, since they focus exclusively on translation. Moreover, they usually lack of a realistic user self-placement mode of introducing sensor displacement and solely rely on a displacement "induced" by the expert. Besides, most of these datasets are proprietary and not freely available for their use. Thus, a dataset that supports real-world sensor displacement is missing. Here an open-access[5] dataset that provides multiple concepts of sensor placement and displacement, namely ideal-placement, self-placement and induced-displacement is presented.

A predefined sensor deployment is habitually considered for activity recognition tasks (e.g., Figure 4.5(a)). The recognition system is usually trained on this ideal setup. However, users may introduce sensor displacement as consequence of the self-placement of the sensors or during the performance of the exercises (e.g., Figure 4.5(b)). Larger displacements may be obtained through an intentional depositioning of the sensors (e.g., Figure 4.5(c)), which may be used to investigate the effects of displacement at worst conditions. To study all these cases the following three scenarios regarding sensor deployment in real-world settings are defined (also summarized in Table 4.1):

- *Ideal-placement* or *default* scenario. The sensors are positioned by the instructor on predefined locations within each body part.

[5]The dataset is accessible at http://www.ugr.es/~oresti/datasets.

Type of displacement	Description
Ideal-placement	No displacement is introduced. This is the default sensor deployment.
Self-placement	The users are asked to position a subset of the sensors by themselves on the specified body part/s (e.g., "please position sensor X on your left thigh").
Induced-displacement	Sensors are intentionally displaced from the ideal position by the expert, typically introducing a large displacement.

Table 4.1: *Methods for implementing realistic sensor displacement.*

The data stemming from this scenario could be considered as the 'training set' for supervised activity recognition systems.

- *Self-placement.* The user is asked to position a subset of the sensors themselves on the body parts specified by the instructor, but without providing any hint on how the sensors must be exactly placed. This scenario is devised to investigate some of the variability that may occur in the day to day usage of an activity recognition system, involving wearable or self-attached sensors. Normally, the self-placement will lead to on-body sensor setups that differ with respect to the ideal-placement. Nevertheless, this difference may be minimal if the subject places the sensor close to the ideal position.

- *Induced-displacement.* An intentional depositioning of sensors using rotations and translations with respect to the ideal placement is introduced by the instructor. One of the key interests of including this last scenario is to investigate how the performance of a certain method degrades as the system drifts far from the ideal setup.

4.4.2. Dataset for displacement evaluation

In the following the particular dataset collected as part of this work is described. The diverse concepts of displacement described in the previous section are here utilized.

(a) (b) (c)

Figure 4.5: *Example of possible sensor placements according to the (**a**) ideal (**b**) self-placement and (**c**) induced-displacement deployments. In (**b**) the sensor is arbitrarily rotated 180° (approx.) by the user with respect to the ideal positioning (**a**). In (**c**) the expert explicitly displaces the sensor from the middle upper arm to the elbow.*

Activity set

The dataset consists of a set of up to 33 typical warm up, fitness and cool down exercises (see Table 4.2). In particular, the dataset includes activities involving translation (L1-L3), jumps (L4-L8) or general fitness exercises (L31-33) as well as body part specific activities focused on trunk (L9-L18), upper extremities (L19-L25) and lower extremities (L26-L30). Diverse reasons support to considering this particular activity set. The activities were selected so that different combinations of body parts are involved in each exercise. Some activities imply the motion of the whole body (e.g., walking or jumping) while others focus on training individual parts (e.g., legs for cycling). Since the activities were very easy to perform, participants had no difficulty in doing the exercises. This simplifies the recording process, helps to collect abundant data and allows for the natural behaving of the users. The exercise type also influences the impact of the displacement on the sensor signals. Rotation related anomalies will be constantly present, even when the sensor remains still, due to the orientation drift measured with respect to the gravitational component for accelerometers. On the other hand translation related anomalies might particularly be observable when the sensor is in motion.

Activity set	
L1: Walking (1 min)	L18: Upper trunk and lower body opposite twist (20x)
L2: Jogging (1 min)	L19: Arms lateral elevation (20x)
L3: Running (1 min)	L20: Arms frontal elevation (20x)
L4: Jump up (20x)	L21: Frontal hand claps (20x)
L5: Jump front & back (20x)	L22: Arms frontal crossing (20x)
L6: Jump sideways (20x)	L23: Shoulders high amplitude rotation (20x)
L7: Jump leg/arms open/closed (20x)	L24: Shoulders low amplitude rotation (20x)
L8: Jump rope (20x)	L25: Arms inner rotation (20x)
L9: Trunk twist (arms outstretched) (20x)	L26: Knees (alternatively) to the breast (20x)
L10: Trunk twist (elbows bended) (20x)	L27: Heels (alternatively) to the backside (20x)
L11: Waist bends forward (20x)	L28: Knees bending (crouching) (20x)
L12: Waist rotation (20x)	L29: Knees (alternatively) bend forward (20x)
L13: Waist bends (reach foot with opposite hand) (20x)	L30: Rotation on the knees (20x)
L14: Reach heels backwards (20x)	L31: Rowing (1 min)
L15: Lateral bend (10x to the left + 10x to the right)	L32: Elliptic bike (1 min)
L16: Lateral bend arm up (10x to the left + 10x to the right)	L33: Cycling (1 min)
L17: Repetitive forward stretching (20x)	

Table 4.2: *Warm up, cool down and fitness exercises considered for the activity set. In brackets the number of repetitions (Nx) or duration of the exercises (in minutes).*

Study setup

A set of nine inertial measurement units (Xsens MTx, [111]) are distributed on the subject's body as shown in Figure 4.6. These nodes provide several sensing modalities including acceleration, rate of turn, magnetic field and derive the orientation estimates of the sensor frame with respect to the Earth reference. Sensors and Xsens Master are wired together in a serial connection that nevertheless do not limit users mobility. The Master device is interfaced over Bluetooth to a laptop which continuously stores the information delivered by the nodes. The laptop is also used for labeling purposes. Both data storage and labeling processes are performed using the CRN Toolbox [112]. The sampling rate is established to 50Hz which suffices for the exercises requirements.

Eight of the sensors are normally positioned on the middle of the limb (for each extremity). An additional one is centered on the back, slightly below the scapulae. The sensors were attached to the body using elastic straps and velcro. Trousers and sports jackets of different sizes were provided in order to ensure the fit to the user.

All sessions were recorded using a video camera. The video recording is useful to check anomalous or unexpected patterns in the data and

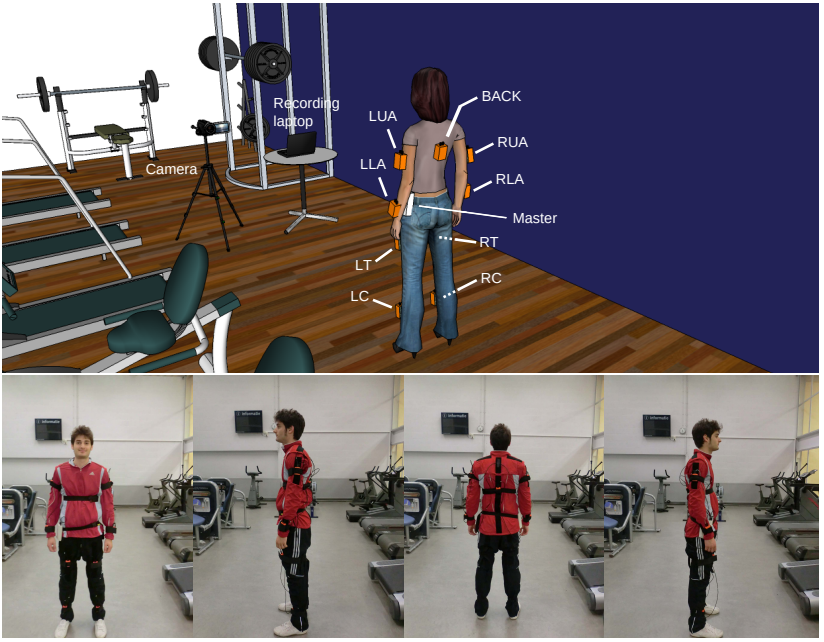


Figure 4.6: *Experimental setup (cardio-fitness room). Eight Xsens units are placed on each body limb and an additional one on the back. A laptop is used to store the recorded data and for labeling tasks. A camera records each session for offline post-processing. Sensor legend: left calf (LC), left thigh (LT), right calf (RC), right thigh (RT), back (BACK), left lower arm (LLA), left upper arm (LUA), right lower arm (RLA), and right upper arm (RUA).*

correct labeling mistakes. In some recordings two subjects performed the exercises in parallel for efficiency.

Experimental protocol

The experiments took place in a cardio-fitness room at the *Student Sport Centre Eindhoven*. The recordings were performed for 17 volunteers, seven females and ten males, with ages ranging from 22 to 37 years old. The experiment consisted in performing two complete runs of the exercises, once with the self-placed and once with the default sensor setup. The self-placed run was performed first as to not give any

clues on the default sensor position to the participant. One run-through of the exercises lasted 15-20 minutes. Each session was preceded by a preparation phase lasting around 30 minutes. For the self-placement recordings, first the users self-positioned three[6] out of the nine sensors, while the remaining sensors were attached by the specialist on their predefined placements. For the second run, self-placed sensors are relocated in their default placements. The preparation phase also comprised the connection of the sensors to the XBUS Master, setting up the video camera, and establishing the Bluetooth connection between the Xsens Master and the laptop. Before starting the exercises, the exact position of the sensors was documented using the video recording.

Three[7] out of the 17 volunteers were recorded for the induced-displacement scenario (concretely subject 2, 5 and 15). For this, the instructor depositioned a subset of the sensors while maintaining the rest in their original location. Data for various sensor configurations were registered, concretely for the case in which four, five, six or even seven out of the nine sensors are misplaced. The participants performed a specific run for each sensor configuration. For both, self-placement and induced-displacement scenarios the displaced sensors were neatly selected to fairly cover all body parts.

An instructor demonstrated each exercise before the user performed them, although the participants were asked to freely execute the activities while trying their best. In general 20 repetitions were recorded for each activity except for those exercises that required the subject's interaction with gym machines (i.e., L1-L3 and L31-L33 from Table 4.2) for which roughly a minute of exercising was recorded. This constitutes an interesting means of gaining a relevant number of instances for each class. The two runs were separated by a break during which the battery levels were checked and the setup for the next run prepared. The exercises were labeled online using the CRN toolbox [112]. The inevitable errors in the online labeling were eliminated in post processing with the help of a commentary sheet to track the errors in addition to the video recording.

[6]This is considered a reasonable estimate of the proportion of sensors that may be misplaced during the normal wearing of the devices.

[7]Conversely to the self-placement case in which the inter-subject variability wants to be observed when self-placing the sensors, this scenario rather focus on the study of the effects of large displacements that are purposely introduced.

Deployment	#subjects	#anomalous sensors	#activities duration/total duration
Ideal-placement	17	0	226.84/860.19
Self-placement	17	3	220.73/895.42
Induced-displacement	3	{4,5,6,7}	{47.72,45.39,48.58,46.76}/632.28

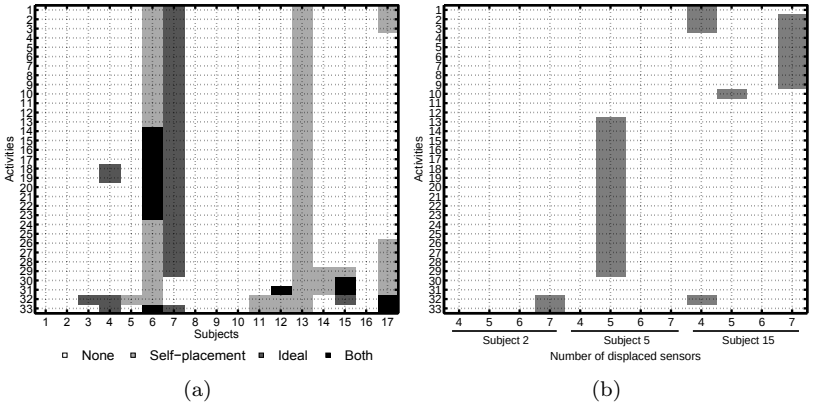
Table 4.3: *Dataset description summary. Overall cumulative duration for the complete set of activities with respect to all the collected data is given in minutes.*

Statistics

Table 4.3 presents the statistics on the available exercise data with respect to the complete recorded data for each displacement concept considered in this work. The entire dataset contains over ten hours of exercise data and lasts over 39 hours in total. Self-placed and ideal recording sessions contain approximately 15 hours of data each. Induced-displacement sessions include more than ten hours of data distributed among the different runs. The difference between exercise duration and the total duration provides the amount of data corresponding to unrelated activities. The average $\pm$ standard deviation duration in minutes of the exercise data (total data) recorded per subject is 13.02 $\pm$ 5.26 (51.31 $\pm$ 20.35) for the ideal concept, 13.96 $\pm$ 3.78 (50.59 $\pm$ 17.41) for the self-placement scenario and 14.75 $\pm$ 5.36 (49.24 $\pm$ 22.43) for the induced case.

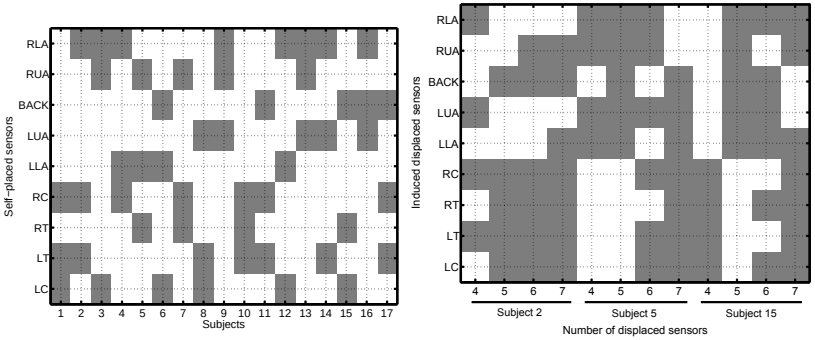
During the data post-analysis some parts of the recordings were identified as either corrupted or missing. The recorded videos were demonstrated especially useful for rejecting erroneous labels as well as checking the validity of the annotated data. Figure 4.7 shows the missing activity data[8] for each subject or run. No activity data is available for subject 6 and 13 for the self-placement setup. For participant 7 there is almost no data available in the ideal scenario. A few additional activities are missing for some of the remaining subjects. For the induced-displacement dataset (Figure 4.7(b)), the worst data loss was incurred for subject 3 where activities L13 to L29 are missing. A few additional activities are also missing for the rest of the subjects. Finally, the sensors that have been displaced for each subject and setup are depicted in Figure 4.8.

[8]The amount of missing data is negligible compared to the number of valid samples, thus it was discarded to redo the corresponding experiments.



(a) (b)

Figure 4.7: *Missing activity data for each particular subject. (**a**) For ideal and self-placement conditions: the legend identifies the corresponding sensor deployment (both ≡ self-placed and ideally-placed). (**b**) For the induced-displaced condition: only participants 2, 5 and 15 were considered.*



(a) (b)

Figure 4.8: *Shading spots identify the displaced sensors for the (**a**) self-placement and (**b**) induced-displacement deployments. Only participants 2, 5 and 15 were considered in (**b**).*

4.4.3. Statistical evaluation of realistic sensor displacement

A preliminary analysis of the variability captured in this dataset is here conducted to underline that it is indeed useful for studying displacement related sensor anomalies and can be considered for benchmarking displacement effects. Mean and standard deviation features are considered for acceleration signals across all sensors. The data is partitioned in such a way that each instance corresponds to roughly one repetition of a given activity.

Figure 4.9 gives a particular example of how the feature distribution is affected by sensor displacement. It can be observed the shift in the feature space for the displaced sensor between the ideal and self-placement scenario. Figure 4.10 shows the variance along the principal components of the features across all subjects for each activity performed for the self-placement and the default scenario. In the ideal sensor placement the variance observed is due to the intra and inter subject variability in performing the exercises. The random sensor displacements introduced in the self-placement scenario result in a higher overall variance, which can be seen by comparing Figure 4.10(b) to Figure 4.10(a). Figures 4.11(a) to 4.11(f) provide a similar illustration for the induced-displacement runs. Here it can be further observed an increasing variance as the number of displaced sensors grows.

An indicator of the average variance for each scenario is shown in Figure 4.12 by marginalizing over the activities and feature dimensions. Figure 4.12(a) compares the average variance for the ideal and self-placement setup. Even though the users only displaced three out of nine sensors within the specified body segment in the self-placement scenario, the increase in variance compared to the ideal setup is considerable. In Figure 4.12(b), a higher variability is once again observed as the number of anomalous sensors increases. These results suggest that activity recognition systems that do not consider sensor displacement mask a huge source of variability.

It can be considered that the activity instances for a given participant form a cluster in the feature space (Figure 4.9). Then, by calculating the normalized cluster distance for a given subject and activity between the ideal and an anomalous setup it is possible to obtain the shift in the feature space caused by the sensor displacement. Figure 4.13 shows the normalized cluster distance between the ideal and self-placement run for subject 10. The feature directions with the highest cluster distances indicate the three particular sensors (LT, RT and RC,

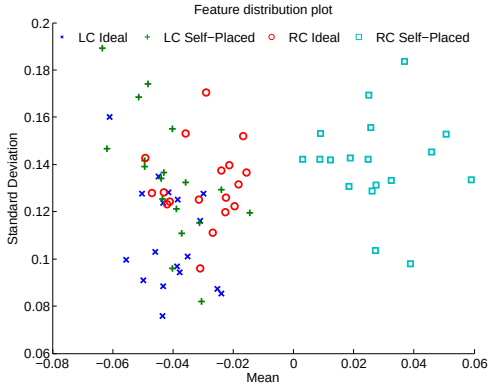
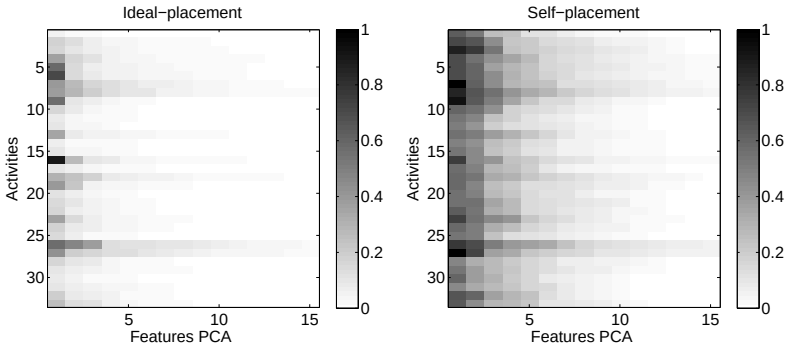


Figure 4.9: *Example of the data drift introduced by the sensor self-placement. The sensor attached to the right calf (RC) was one of the sensors positioned by the user during the self-placement recording session while the left calf (LC) was placed in the ideal position. A shift in the data distribution may be observed between ideal and self-placement run for the displaced sensor, while that of the non-displaced sensor stays the same.*



(a) (b)

Figure 4.10: *Normalized variance across participants for each activity along the 15 most significant principal components for the ideal and self-placement scenarios. The components are evaluated over the mean and standard deviation features extracted from the tri-axial accelerometer measurements for the nine sensors.*

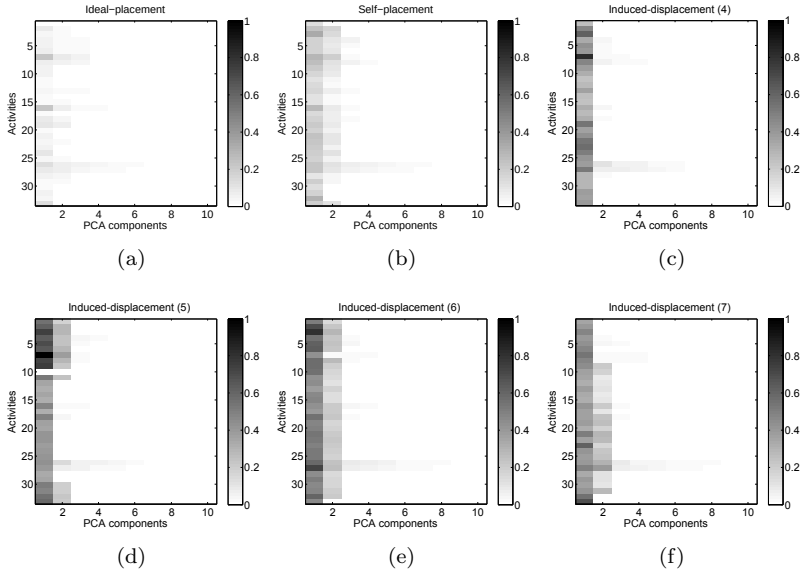
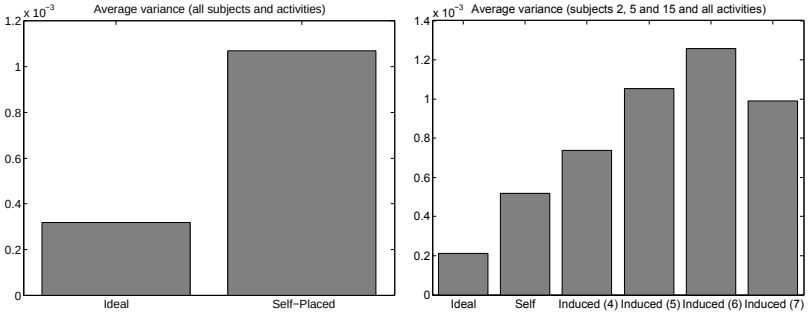


Figure 4.11: *Normalized variance across participants 2, 5 and 15, for each activity along the ten most significant principal components for the ideal-placement, self-placement and induced-displacement scenarios. The components are assessed over the mean and standard deviation features extracted from the tri-axial accelerometer measurements for the nine sensors.*

see Figure 4.8) the participant displaced as a consequence of their self-attachment. Highest variations are observed in the acceleration mean, which especially relates to the gravitational component.

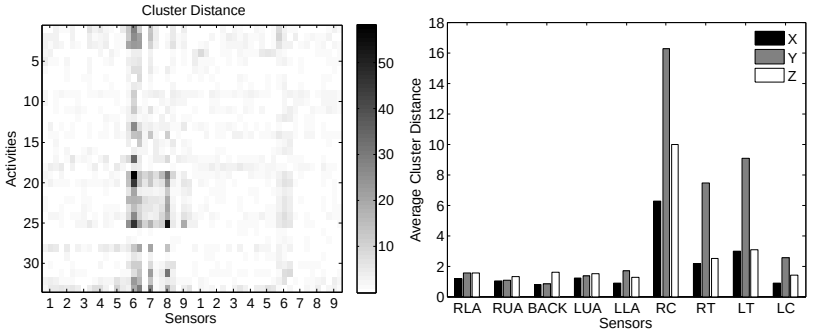
4.4.4. Classification impact of realistic sensor displacement

This section analyzes the tolerance of AR systems to the effects of sensor displacement measured in a realistic settings. Similarly as to Section 4.3, here both single and multiple sensor configurations are evaluated.



(a) (b)

Figure 4.12: *Variance across subjects averaged over activities and features. In (**b**) the ideal and self-placement results are obtained from the data of participants 2, 5 and 15. For the induced-displacement the number of depositioned sensors is given in brackets.*



(a) (b)

Figure 4.13: *Example of the cluster distances between activities for ideal and self-placement recordings for participant 10. Sensors RC, RT and LT were self-placed. In (**a**), clusters were obtained from mean and standard deviation features along X, Y, and Z axes of the acceleration measurement of each sensor (1 to 9 ≡ RLA, RUA, BACK, LUA, LLA, RC, RT, LT, LC) for all annotated dataset activities. In (**b**) only the mean is considered along the same tri-axial directions.*

Experimental setup

Taking advantage of the amount of exercises registered in this dataset, diverse scenarios are defined to analyze the impact of the activity recognition problem complexity on the robustness of the systems to sensor displacement. Two reduced versions of the original dataset, i.e., comprising a subset of the whole group of activities, and the actual original dataset are used for evaluation. Concretely, these datasets respectively embrace 10-activities, 20-activities and 33-activities (all) from the original set. To ensure a fair distribution of the diverse type of activities registered for this dataset, exercises that involve the motion of different parts of the users' body are selected for the 10-activities and 20-activities datasets. The selected exercises are *1*, *4*, *8*, *10*, *12*, *18*, *22*, *25*, *28*, and *33* for the 10-activities scenario, and *1*, *2*, *3*, *7*, *12*, *13*, *17*, *18*, *19*, *20*, *21*, *23*, *25*, *27*, *28*, *29*, *30*, *31*, *32*, and *33* for the 20-activities case (see Table 4.2 for equivalence). The data corresponding to these activities are chosen for the diverse concepts of sensor placement and displacement ('ideal', 'self' and 'induced').

Here again, AR systems based on a single or individual sensor and those based on a multi-sensor configuration are respectively tested[9]. For the single sensor approach a sole ARC is required (SARC). For the multi-sensor system, FFMARC and HWC models are approached. A segmentation process consisting of a non-overlapping sliding window (6 seconds size, as suggested in [10]) is applied to each data stream. Three feature sets (FS) are respectively extracted for evaluation: FS1='mean', FS2='mean and standard deviation' and FS3='mean, standard deviation, maximum, minimum and mean crossing rate'. These are features widely used in activity recognition [43, 10, 16, 5, 39] for their discrimination potential and ease of interpretation in the acceleration domain. Additionally, the use of these features may also simplify the task of reproducing these experiments for future work comparison. Likewise, three of the most extensively and successfully machine learning techniques used in previous activity recognition problems are considered for classification: C4.5 decision trees (DT, [102]), k-nearest neighbors (KNN, [103]) and naive Bayes (NB, [95]). The k-value for the KNN model is particularly set to three as it has been shown to provide good

[9]All the processing is performed in Matlab R2011b. For the preprocessing, featuring and classification stages some of the functions provided in the *Signal Processing*, *Statistics* and *Bioinformatics* toolboxes have been used, while many others have been specifically defined for this purpose.

results in prior work of this thesis. These techniques are used for the standard classifiers and for the HWC base classifiers units.

The evaluation of the activity recognition models for the 'ideal' settings is performed through a ten-fold random-partitioning cross-validation process applied across all subjects and activities. This provides the baseline recognition capabilities of these systems in absence of displacement. This process is repeated 100 times for each method to ensure statistical robustness. A testing procedure is applied for 'self' and 'induced' settings. Thus, self-placement and mutual-displacement data is inputted to a predefined AR system which is obtained from the training on the data registered for the ideal-placement setting.

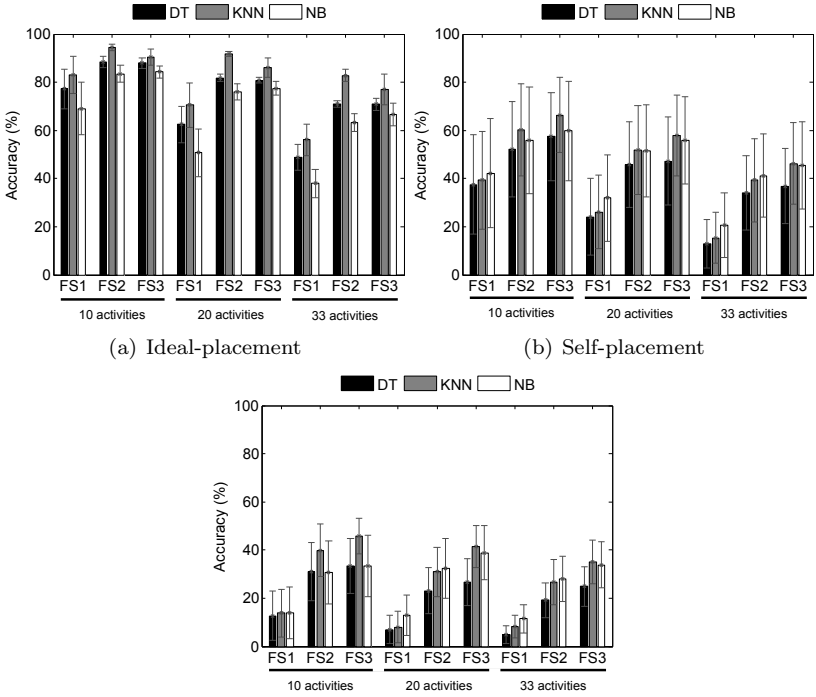
Single sensor performance

In this section the results corresponding to the evaluation of the SARC systems are presented. The performance results obtained from the evaluation across each individual sensor are depicted for each displacement-concept setting in Figure 4.14.

In general terms, two tendencies could be discerned for all settings and scenarios. First, as the complexity of the problem increases, i.e., the number of considered activities grows the detection performance reduces. Second, the use of richer feature sets helps improve the recognition capabilities of the systems, thus best results are normally obtained for FS3, followed by FS2 and FS1. This is consistent with previous studies of this dissertation.

For the ideal-placement (Figure 4.14(a)), from all evaluated models the KNN stands out with values above 90% accuracy for the 10-activities and 20-activities problems. The performance drops to a bit more than 80% for the 33 activities scenario. The other two paradigms demonstrate limited applicability for recognition, specially for the most complex scenarios.

A significant drop on the recognition capacity is seen when the users self-place the sensors (Figure 4.14(b)). Through comparing these results with the ones obtained for the ideal case, for the simplest scenario (i.e., 10-activities) a drop of approximately 40% accuracy is observed for the KNN approach and FS1, that nevertheless gets reduced to almost 35% and 25% drop for FS2 and FS3 respectively. This "enhancement" for FS2 and FS3 could be explained since the use of richer feature vectors may help compensate the sensitivity to displacement variations of the features used in FS1. The performance gap is higher for the



(a) Ideal-placement

(b) Self-placement

(c) Induced-displacement

Figure 4.14: *Accuracy (average - bar - and standard deviation - whiskers -) results from the evaluation of the* ***single sensor*** *approach across all subjects and sensors for the (**a**) ideal-placement, (**b**) self-placement and (**c**) induced-displacement settings. Top legend identifies the classification paradigm. Horizontal axis labels identify the feature set used for each experiment. The AR dataset (i.e., number of activities) used is respectively underlined.*

20-activities and 30-activities scenarios, on which the accuracy falls up to 45%. The tendencies observed for the KNN are similarly applicable to the DT model. On the other hand, NB appears to be the most robust approach. The drops are very similar for all scenarios (around 20-25%, the lowest of evaluated). In fact, the performance for NB after displacement is quite similar to as for the KNN model. In either case, since the maximum performance across all models and paradigms is of

65%, it can be concluded that none of the models satisfactorily cope with the displacement introduced by the users when self-placing the sensors. In other words, these models are not practical for recognition purposes under these circumstances.

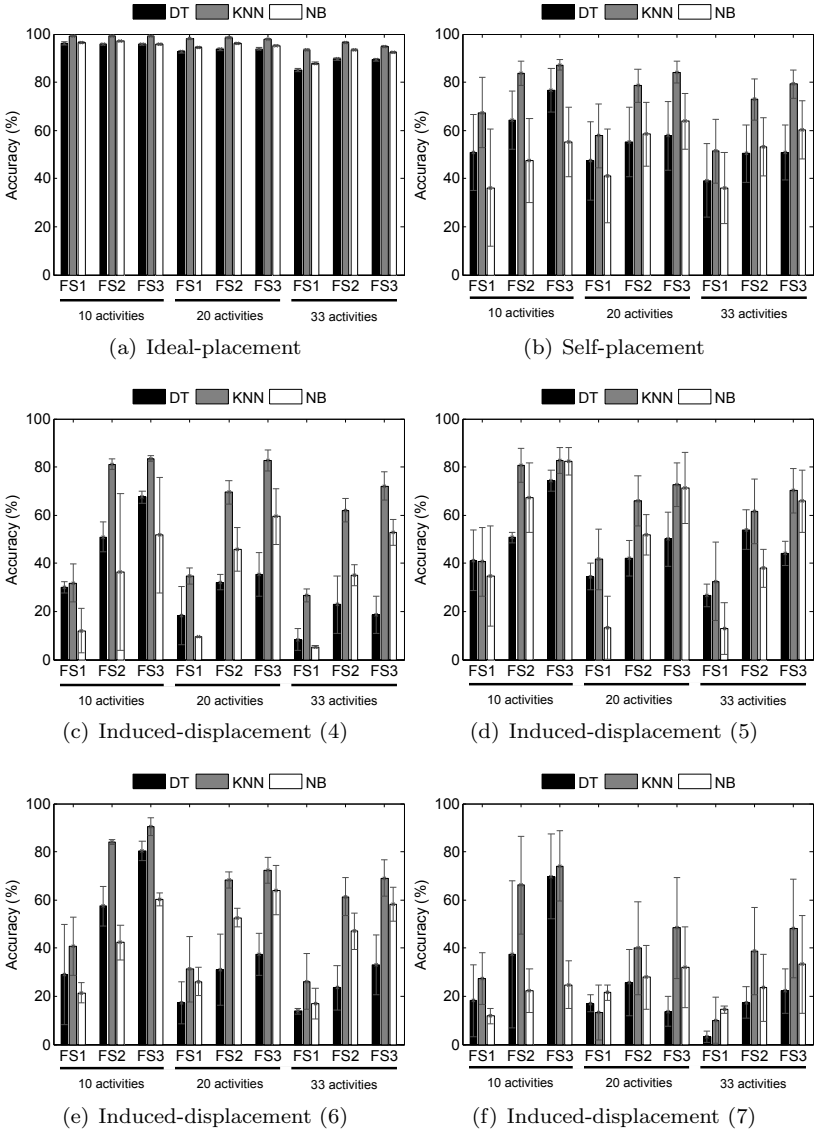
Finally, for the induced-displacement setting, the data corresponding to the displaced sensors of all categories (four, five, six and seven displaced sensors) are used for evaluation. Similarly as to for the self-placement setting, an overall analysis of the effects for each sensor (i.e., body location) is here feasible since most possible combinations of displaced sensors where considered during the dataset recording process (see Figure 4.8). From Figure 4.14(c) it could be seen the tremendous performance worsening that the diverse recognition systems suffer under the assumption of an intentional depositioning of the sensors. Clearly, these are the poorest results among the three settings with performance values that are below 50% accuracy at best. As for the self-placement setting, the minimum performance fall is encountered for the NB model, while the highest drop is found for the KNN technique, although this yields best performance values from the evaluated models.

Multi-sensor fusion performance

The accuracy results for both feature fusion (FFMARC) and decision fusion (HWC) approaches are respectively depicted in Figures 4.15-4.16. It should be noted that differently to the individual sensor analysis here all sensors are simultaneously used, then the evaluation could be respectively performed on the diverse categories devised for the mutual-displacement setting (i.e, four, five, six or seven out of the nine sensors are depositioned).

Similarly to what was commented for the individual sensor approach, it could be seen that the performance of the diverse recognition systems improves as the number of features grows. On the other hand, the performance decreases as the activity detection problem gets more complex (i.e., number of activities increases), albeit this performance fall is less marked than for the SARC models.

The feature fusion model demonstrates high recognition capabilities for all AR scenarios when an ideal placement of the sensors is considered (Figure 4.15(a)). In fact, the performance is near absolute recognition for models implementing the KNN technique, with an accuracy of more than 95% for the most complex scenario. Also very high figures are



(a) Ideal-placement

(b) Self-placement

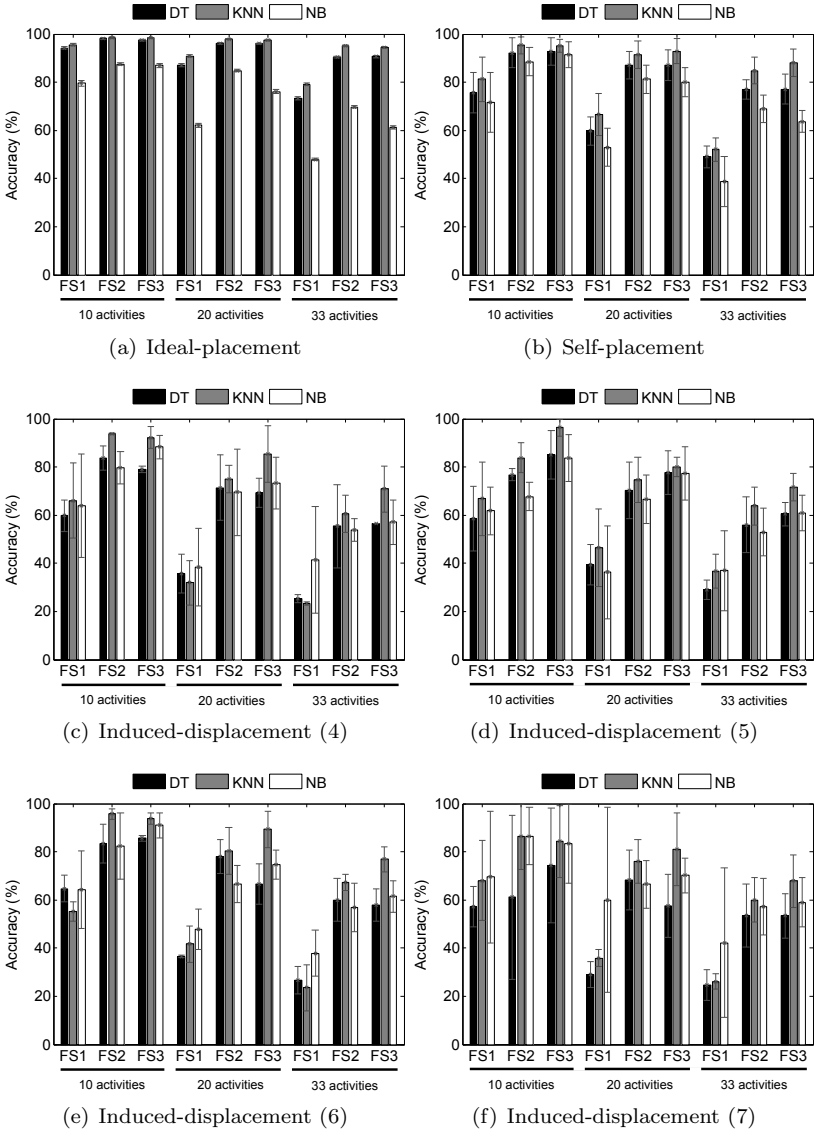
(c) Induced-displacement (4)

(d) Induced-displacement (5)

(e) Induced-displacement (6)

(f) Induced-displacement (7)

Figure 4.15: *Accuracy (mean and standard deviation) results from the evaluation of the* ***feature fusion*** *model for the (**a**) ideal-placement, (**b**) self-placement and (**c-f**) induced-displacement (# sensors) settings. Top legend identifies the classification paradigm. Horizontal axis labels identify the feature set used for each experiment. The AR dataset (i.e., number of activities) used is respectively underlined.*



(a) Ideal-placement

(b) Self-placement

(c) Induced-displacement (4)

(d) Induced-displacement (5)

(e) Induced-displacement (6)

(f) Induced-displacement (7)

Figure 4.16: *Accuracy (mean and standard deviation) results from the evaluation of the* ***decision fusion*** *model for the (**a**) ideal-placement, (**b**) self-placement and (**c-f**) induced-displacement (# sensors) settings. Top legend identifies the base classifiers paradigm. Horizontal axis labels identify the feature set used for each experiment. The AR dataset (i.e., number of activities) used is respectively underlined.*

obtained for the other two methodologies. DT proves to be the less reliable from all methods, yet it provides performance values above 90% for the 10 and 20-activities scenarios. Finally, NB yields accuracies that spans from 87% for the most complex scenario and FS1, to 97% for the simplest scenario and FS2 or FS3.

Different readings could be extract for the self-placement case (Figure 4.15(b)). Here the KNN model stands out as the most reliable technique and also the most robust as this suffers from the lowest performance falls with respect to the default setup. Despite there is a moderate drop for the simplest AR scenarios when rich feature sets are used (84% and 87% accuracy for the 10 and 20-activities while using FS3), a considerable drop is found for the most complex scenario that leads to average top performances of less than 80%. Highest drops could be seen for the use of FS1 as feature input, with average performance values that reach 70% accuracy at best and for the simplest scenario, to almost 50% for the 33 actitivies case. In summary, the performance fall approximately spans from 13% at best to more than 40% at worst case. The worsening is even higher for the DT and NB, with performance drops that range between 20% and 60%.

The effects of the induced-displacement are shown to be in general more harmful than for the self-placement case. Similarly as for the ideal and self-placement cases, KNN proves to be the most accurate model in general terms. Nevertheless the performance falls differently for each scenario. Starting with the case in which four sensors are intentionally depositioned (Figure 4.15(c)), it can be observed that a maximum accuracy above 80% is guaranteed for FS3 and 10 or 20-activities. This represents a reduction of approximately 20% with respect to the ideal case. For the rest of evaluations the drop is very significant, falling in some cases up to 90% with respect to ideal circumstances. When five sensors are displaced the performance drops span from 20% to 85% at worst. An improvement is observed for the case of depositioning six sensors for the 10-activities scenario, although the performance worsens for the other two scenarios (up to 75% loss). Finally, as could be expected worst results are obtained when seven out of the nine sensors are displaced. In average terms, none of the models gets a performance superior to 75%, specially for the most complex scenarios, on which the highest drop is about 85%. To explain the singularity found for the six-sensors-displaced setting (better overall results for the simplest scenario than for four and five) it must be born in mind that for the induced-displacement evaluation the recordings for just three participants are

available, thus the conclusions could be in principle less generalizable than for the self-placement case. Moreover some data are missing for part of the users as illustrated in Figure 4.7(b), particularly for one of the volunteers that participated in the induced-displacement (5) recordings, which may also explain these trends. The differences among these could be also explained through the random nature of the displacement introduced from run to run, thus larger sensor displacements may be present for this settings than for others. In either case, this idea fits well with what can be observed in realistic settings.

According to the HWC model, in ideal settings very promising results are obtained for all scenarios, although this specially applies to DT and KNN models. Accuracy rates above 90% and 95% may be respectively obtained for DT and KNN models for all scenarios and features sets, albeit these reduce for the most complex scenarios when the simplest feature set is used (FS1). The NB technique proves to be the less reliable, with performance values that improve when richer feature sets are used. Although performances over 87% and 85% could be obtained for the 10 and 20-activities scenarios, only a 70% accuracy may be achieved for the 33-activities case when using the NB technique.

The HWC model significantly overcomes the worsening experienced by SARC and FFMARC models when the user self-places the sensors, specially for the most complex scenarios. Similarly as for the feature fusion model, KNN is once again the most promising technique to be used as base classifier. For this, when FS2 and FS3 are used the performance achieved is superior to 95% for the 10-activities scenario, 92% for the 20-activities case and close to 90% for the 33 activities scenario and FS3. This translates into very reduced performance drops that span from 3% to 6% worst. As it happens to occur to SARC and FFMARC models, the approaches that utilize the FS1 are not capable of dealing with the effects of displacement, however, the drops are yet lesser than for the single sensor and feature fusion models. DT and NB models appear to be quite robust considering their moderate performance drops, which in most cases are as reduced as for the KNN approach. This clearly supports that such robustness falls on the HWC structure, since the behavior is quite similar with independence of the machine learning paradigm used for the base units.

Similarly as in the previous cases, the combination of HWC and KNN turns to be the most reliable model under the assumption of induced displacements (Figure 4.16(c)-4.16(f)). Besides, NB confirms as the approach that, even not yielding best average performance, suffers

from the lowest falls. From an individualized observation, performance rates of up to 94% may be achieved for the simplest scenario when four sensors are depositioned. This performance decays to 85% and 70% for the 20-activities and 33 activities scenarios, quite in line with the results obtained for the feature fusion, which here nevertheless correspond to a lower performance fall. Concretely, the performance drops for this case range from 5% to 35% for KNN and FS2 or FS3, and reach a 60% drop when FS1 is used. For NB and DT the falls are found to be even lesser, with a performance improvement only shown for a few cases. The evaluation on the data pertaining to the depositioning of five sensors gives results that show no deterioration of the performance for KNN-FS3 and 10-activities, while a significant worsening is seen for the most complex scenarios. Particularly, the performance falls range between 2% and 45% for this setting. Slightly better results are obtained for the six-sensors-displaced case, in which accuracies of more than 90% are reached for the 10 and 20-activities scenarios, while little more than 75% could be achieved for the 33 activities problem. In overall, the performance drop spans from 3% at best to 55% at worst case. Finally, worst results are obtained for the seven depositioned sensors setting. Newly, NB proves to be the most robust technique whereas maximum performances are generally obtained for KNN. The performance drop ranges from 2% to more than 55%. The high standard deviation values for this last setup reflect the differences about the deployment used for each user and run.

4.4.5. Discussion

In general terms, two tendencies could be discerned for all results independently of the sensor deployment settings. First, as the complexity of the problem increases, i.e., the number of considered activities grows, the detection performance reduces. Second, the use of richer feature sets helps to improve the recognition capabilities of the considered systems, thus best results are normally obtained for FS3, followed by FS2 and FS1. This is consistent with previous studies of this dissertation.

The accuracy results demonstrate that the evaluated AR approaches could be adequate solutions under the assumption of an invariant sensor deployment. Nevertheless, both sensors self-placement and induced-displacement settings introduce a significant drop on the performance of standard AR systems with respect to an ideal sensor deployment, which confirms the effects described in section 4.2. Clearly, the more

profound the displacement applied the higher the performance drop. Likewise, the more sensors are depositioned in a multi-sensor setup, the more limited the recognition accuracy is. The HWC model proves to be the most robust approach from all evaluated systems, capable of dealing with the variations introduced during the user self-placement of the sensors and also overcoming the impact of large depositionings for particular scenarios.

Single vs. multiple sensing

For the single sensor-based approach, KNN appears to be the most reliable technique in ideal circumstances, but possibly the most sensitive to sensor displacement. This could be consequence of the way the activity clusters drift in the feature space because of the displacement effects. On the contrary, models based on NB appear to be the less accurate for predefined conditions, although these suffer from the lowest drop on the performance when the sensors are misplaced. In any case, the maximum performance obtainable in the event of sensor displacement do not suffice for the purpose of AR, thus once again demonstrating the potential limitations of SARC approaches in realistic settings.

Models based on a multi-sensor setup generally allow for an optimal recognition capabilities given a default deployment of the sensors. Besides, more promising results are obtained under the effects of sensor displacement than for the individual sensor approach. Nevertheless, not all models demonstrate the same tolerance to sensor misplacement. In fact, the FFMARC model shows an important worsening for the case in which the user self-places the sensors. The performance drop is even higher when the sensors are purposely depositioned. For the FFMARC model NB proves to be the most sensitive to sensor drifts, while the opposite is observed for the KNN technique.

On the contrary, the HWC model deals quite well with the effects of sensor displacement introduced by a user self-placement of the sensors. The HWC model also demonstrates a high tolerance to large depositionings for simple AR scenarios. This is shown to happen independently of the classification paradigm utilized for the base classifiers. However, as the number of displaced sensors increases, the performance of both FFMARC and HWC approaches generally declines. The recognition capabilities are specially reduced when a majority of the sensors are displaced and for very complex recognition problems. Even though, comparatively here again systems based on the HWC scheme signifi-

cantly overcome the robustness capacity of FFMARC models.

HWC advantages

The HWC model provides us with several benefits some of which have been already described in previous sections of this dissertation. Some of them are further supported by the results obtained in this study, while new others are specifically identified from the analysis of realistic sensor displacements.

HWC assures a higher robustness to sensor displacement effects than standard AR approaches. These better results for the HWC may be explained since individual variations of a sensor with respect to its default behavior have less impact in the classification process. This is possible since each sensor contributes in an independent manner to the final delivered decision, thus a majority of sensors (normally unaffected) overcomes the decisions obtained from the displaced sensors. Conversely, single sensor models and feature fusion models incorporate data from displaced sensors in a unique feature vector, thus leading to a potential feature drift that cannot be handled by the reasoning model.

The HWC model demonstrates an important tolerance to sensor displacement, although it is subject to a certain performance drop. However, it is worth noting that this drop applies homogeneously to each classification paradigm, which leads us to conclude that the potential robustness of the system truly relies on the HWC structure, more than the specific techniques used for the base classifiers. In this way, the HWC model proves to increase the classification potential and robustness of standard multi-class models.

Although a comparison with previous works could be performed for the synthetic displacement analysis (Section 4.3.5), here this is not possible due to the novelty of this approach. To the best of our knowledge there is no investigation that treated on-body sensor displacement in this regard, thus this contribution could be thought as a starting point of investigation in this field.

Finally, the HWC demonstrates high potential for its application in ideal circumstances. Moreover, the proposed model also shows notable scalability properties. It was shown to provide good performance for reduced sensor networks such as the considered in Chapter 3, and here also successful for more complex deployments. Likewise, the dual weighting scheme (insertion/rejection) also demonstrates of relevant capacity. For the most challenging scenario, the decisions of up to 297

(9x33) base classifiers must be combined, thus an efficient aggregation mechanism is required. The weighting model used seems to operate well in such a dense ecosystem.

Concepts of displacement

This is the first study in this field that widely accounts for the problem of sensor self-placement in the investigation of real-world AR issues. As demonstrated through this work, there are serious consequences when freeing users to place the sensors. AR systems based on a single sensor may be useless if the user does not follow the specific deployment instructions. This is a limitation that goes against usability requirements of real-world AR systems. Sensors are devised to be embedded in articles of the daily living, thus restricting the way these must be worn proves burdensome and could reduce users' acceptance. When porting multiple sensors, the effect of isolated sensor misplacements may be reduced. In fact, it could be seen that a bracelet is worn on the upper arm instead of the lower arm, but it is normally expected that a majority of other instrumented items such as shoes, trousers or shirts are put on a quite predefined manner. In this direction, the HWC model helps overcome the limitations introduced by a minority of drifted sensors.

The induced-displacement setting was originally planned to analyze the effects of a hard depositioning of the sensors. This could be also part of the day to day usage of the systems, but it is rather expected to be seen in occasional circumstances. In real-world settings, this could apply to sensor displacements such as those introduced when rolling up the sleeves of an instrumented shirt or when wearing the shirt back-to-front. In wearable AR, this type of displacement has been discreetly studied but for a particular body part and in a very limited way in terms of number of users and activities. In this work several configurations and settings are analyzed, thus bringing added value with respect to the state-of-the-art. This study also shows the increasing worsening of the recognition capabilities as the number of displaced sensors increases. Neither individual sensor approaches nor feature fusion models are found to be capable of overcoming the challenge of multiple displaced sensors. The HWC model partially deals with this, but for the simplest scenarios.

AR problem complexity

In general, as the complexity of the recognition problem grows an accurate recognition of the diverse activities turns to be more difficult. Something similar is observed under the assumption of sensor displacement. Concretely, for the ten activities scenario low performance reductions are seen, specially when the HWC and richest feature sets are utilized. However, as the number of activities grows also the effects of displacement are more prominent, and by extension also the recognition worsening more profound. This may be explained since the probability of misclassification of a given instance is higher because there are more potential activity clusters with which this may be confused. The use of more features or kernel-type transformations of the feature space could be of interest to separate as much as possible the diverse class clusters and therefore reduce the activity confusion likelihood.

Study generalization

One of the key elements that supports the generalization of the conclusions of this study is the data employed for evaluation. The considered sensor deployment comprises some of the most widely used placements in wearable sensing. Sensors are located in unobtrusive and comfortable positions that permit the normal behaving of users during their daily living. Moreover, the deployment almost covers the complete body, concretely registering the movements of all limbs and trunk. To be perfectly complete, sensors on the head should be mounted, however, this option was dismissed considering its rare application normally motivated to avoid users' discomfort. In terms of activities, a broad number of exercises of diverse type are considered. This includes exercises of different intensity and velocity that also involve diverse combinations of body parts. This helps to get insights into the effects of sensor displacement for activities in which some body parts are still or quasi-static, to others in which these are in motion, thus tackling a wide spectrum of activities. A considerable number of participants were also considered for the recordings, thus better supporting the validation of the study outcomes. Although participants of relatively similar age were considered for evaluation, their physical conditions were quite diverse thus ensuring different kinds of executions. This is further supported through freeing the users to execute the activities in a natural way.

For the sake of generalization, the here tested recognition systems correspond with some of the most widely used in related works. More-

over, simplicity and comprehensiveness were key elements born in mind during the selection of these models, thereby allowing us to focus on the potential impact of the displacement effects. Thus for example, data directly captured through the sensors are used, avoiding any kind of filtering or preprocessing. These procedures normally remove some parts of the raw signals that may potentially lead to a change in the signal space that could mask the actual effects of sensor displacement. Moreover, the features used are simple, easy to calculate and with interpretable physical meaning. Concretely, the 'mean' allows us to extract the gravitational component contribution to the acceleration which is particularly informative for distinguishing among low intensity activities. 'Standard deviation', 'minimum' and 'maximum' provide insights into the intensity and magnitude of the movements, while the 'mean crossing rate' correlates with the dynamicity and frequency of the executions. Moreover, the use of lightweight features helps reduce latency which is a key factor during online operation of the systems. Similar tendencies have been found for the various feature sets for each independent classification methodology, thus demonstrating that the results obtained here could be extrapolated to other systems of similar nature. In either case, the differences among performance quality for each feature set determine that an automatic selection of better features as performed in previous experiments could possibly lead to improved results.

Open issues

The aim of this work was to analyze the effect of static displacements normally introduced during sensors placement or relocations. These displacements remains in principle static during the use of the systems, although small variations of the sensor positioning are sometimes observed because of loose-fitting attachments or when performing the activities. This is in this work leveraged to further extend some of the conclusions to the case of dynamic displacements. Nevertheless, a profound investigation in this direction is required to better determine the specific effects of sensor dynamic displacements.

Future work may also include the evaluation of the effects of extreme sensor depositionings. Extreme sensor depositioning here refers to the displacement introduced when two or more sensors are exchanged, thus relocated in body parts completely unrelated to the devised at design time. For example, it could be expected that a user misplaces a sen-

sor during the self-placement process, consequence of a mistake (e.g., the wrist-sensor is positioned on the ankle). Although less frequent, real-world AR systems should account for these extreme sensor misplacements, thus contributing to a freer use of the devices.

Similarly, for the self-placement setting it would be also interesting to provide total freedom to users to place the sensors wherever they prefer. Although no hints were given to the participants on how they had to put the sensors on, indications about the particular body part were provided. This was necessary to complete the sensor deployment with the default placement of the remaining sensors. Therefore, a study on which the users self-attach the complete set of sensors in an arbitrary manner could be of utility.

When a majority of the sensors are depositioned, and especially for complex scenarios, the proposed models are not capable to deal with the effects of the displacement. Some possible next steps in this direction were already presented in Section 4.3.5. Apart from these, other approaches are here suggested as future work. Considering the structure of the HWC model, an interesting approach would be to automatically update the parameters of the model to reduce the impact of the decisions yielded by the displaced sensors. To that end, ascertaining which sensors have been displaced is completely necessary. This could be performed through the use of the collectivity to identify which sensors are anomalously behaving or through an statistical analysis of the variations of each sensor data stream.

4.5. Conclusions

Most AR systems assume a predefined sensor deployment that further remains unchanged during runtime. However, these are not lifelike assumptions. Restricting the way on how users must place their wearable sensor devices is unrealistic, unpractical and contributes to people's lack of interest in the use of these systems. In fact, when considered these sensors to potentially be embedded in clothes, garments or other portable accessories of the daily living, it must be seriously taken into account the casualness and naturality with which users normally put on these items.

When a sensor is placed in a different position with respect to its ideal or default placement, an effective displacement could be identified from the former to the latter position. This sensor displacement, that could be categorized as a combination of rotations and translations,

normally translates into a drift of the original signal space, variation that further propagates along the complete ARC. This may potentially lead to a malfunctioning of AR systems designed to operate on a definite sensor deployment. Given the importance of this issue, and the shortage of knowledge to this respect, in this chapter a wide analysis of the effects of sensor displacement for diverse setups, methodologies and settings has been elaborated.

To analyze the effects of sensors displacement, systems originally planned for a predefined sensor deployment are tested on data measured after displacement, here demonstrating an affective worsening of the system recognition capabilities as consequence of the displacement effects. In this investigation two main approaches were followed.

Firstly, the effect of displacements are analyzed for motion data synthetically treated to incorporate the effects of rotations and translations. Systems trained under the consideration that the devices remain unaltered during runtime have been demonstrated not to suffice when based on a single sensor (SARC). The comparison of the robustness of sensors located on different body parts demonstrates that those locations which are more constrained in terms of mobility (such as the hip or thigh) are more sensitive to the sensor displacement anomalies.

The use of multi-sensor configurations help compensate the effects of sensor position variations. Feature fusion (FFMARC) and decision fusion (HWC) models are evaluated, yielding different performances. Changes in any of the individual sensors are assimilated in the aggregated feature vector similarly as for the single sensor approach, thus the FFMARC model proves to be too much sensitive but for small displacements. Conversely, and even when the sensor deployment is hardly modified with respect to the originally considered, HWC enhances the robustness of individual sensors in more than 60% at worst conditions when a minority of the sensors are affected. Furthermore, when slight to moderate variations are considered, this decision fusion model copes with the effect of the disturbances independently of the number of affected sensors.

The other investigated approach consists in the observation of sensor displacements in realistic settings. This research focuses on the effects of the displacements introduced during the normal use of the sensors, particularly when wearing the devices. To that end, a novel open-access benchmark dataset was collected as part of this work, to date missing for the study of this problem. A concept for categorizing inertial sensor displacement conditions in ideal-placement, self-

placement, and induced-displacement is introduced. Ideal-placement settings correspond to the case in which the sensors are positioned according to the default sensor deployment devised during the design phase, thus representing a recognition baseline for default conditions. Self-placement settings reflect a user-centered observation of how people place the sensors by themselves, e.g., in a sports or lifestyle application. Induced-displacement conditions correspond to extreme displacement variants and thus could represent boundary conditions for recognition algorithms. The collected dataset is not only devised for the sake of this investigation but may become a valuable tool to compare performance of different methods and conditions. Moreover, with the large set of annotated activities, the dataset lend itself primarily for activity classification problems. A wide variety of sensors were considered for displacements to capture potential effects on a recognition methods feature extraction and recognition.

First, an statistical analysis of the effects of displacement is performed. This evaluation confirms that mean shifts and increased variance can be observed from ideal to self-placement condition. This result is further confirmed by PCA analysis of all acceleration sensors across the activities. Moreover, a cluster-based analysis demonstrates that a substantial distance increase between ideal and self-placement conditions can be obtained. The disparity is even higher for the case in which the sensors are largely depositioned, here observed through the induced-displacement case.

From the diverse concepts of displacement proposed as part of this work, it could be seen that the way users place the sensors in a natural manner could have serious consequences on the systems recognition capabilities. A practical worsening is observed in most recognition systems, specially on those based on a single sensor unit. Actually, systems of an individual sensor seem to be useless under the assumption of displacement. The use of multi-sensor configurations help counteract the variations introduced in the sensor deployment. Nevertheless, not all MARC models demonstrate the same tolerance to sensor displacement. In fact, the performance of feature fusion approaches is subject to a severe worsening for the case in which the user self-places the sensors. The performance drop is even higher when the sensors are purposely depositioned. FFMARC models severely suffer from the effects of displacement since changes in the signal space are naturally incorporated into the aggregated feature vector. Neither SARC nor FFMARC approaches demonstrate acceptable recognition capabilities

after displacement. On the contrary, and here again, the HWC model outstands as the most robust approach from evaluated. Decision fusion models play an interesting role to help avoiding the effects introduced by each individual sensor displacement. This is possible because decisions are considered independently and the collectivity copes with the errors introduced by displaced sensors. The performance drop is more significant when a majority of sensors are displaced in a heavy fashion. In this case, only the HWC model demonstrates accurate recognition but for the simplest AR scenario. In either case, when a majority of the sensors are displaced more robust alternatives should be considered, techniques that are planned as future work of this dissertation.

5

Supporting AR systems network changes: instruction of newcomer sensors

5.1. Introduction

Nowadays technology progresses in a way never thought before. This technological evolution is seen to empower applications with extraordinary new characteristics as well as remarkably improve user experience. To support that, apps are continuously updated, devices are under timely maintenance and systems frequently upgraded. Likewise, AR systems of the real-world requires a constant adaptation to ensure a seamless, efficient and lifelong usage.

Sensing technology is subject to failures or faults that may be irrecoverable. Although models are devised in this work to overcome the effects of sensor technological anomalies, a replacement of the affected sensor may be required to restore the system to full operating capacity. To that end, the damaged sensor is ideally substituted through a new device of similar characteristics. Moreover, not only sensors are replaced when detected not to function but also when higher efficiency or new features are sought. Thus for example, a new sensor that reduces energy needs or proves more robust to sensor failures may replace an early sensor used for the same purpose. Old-fashioned or legacy devices could be also substituted by novel models that fit better with usability and fashionability requirements. For all these cases, the newcomer sensor may potentially have different characteristics to the substituted one (e.g., different sampling rate, dynamic range, modality), then predefined AR systems may be likely incapable of profiting from data obtained through the new sensor.

Sensors could also be newly incorporated to the default sensing topology. The addition of new sensors could be part of a specific system upgrading, e.g., to enhance the recognition accuracy or provide network redundancy. More habitually, users may acquire new gadgets or devices to benefit from other services not supported by the current employed systems. This poses a new sensor setup configuration that is hardly foreseeable during the design phase. Consequently, default AR systems may not directly leverage the data obtained through the novel sensor devices.

Sensors may just happen to be discovered as available to the current user context. In fact, there is a tendency towards an increased availability of sensors readily deployed by users by themselves (e.g., smartphones, sensor-equipped gadgets, smart objects, smart clothing) or integrated as part of living environments (e.g., sensors for climate control, security, or entertainment). In the general case, many of these

sensors may not have associated activity models to use them for activity recognition, as they are deployed for other purposes. However, most of this sensing equipment could be used for AR purposes since they are in principle capable of measuring human behavior (e.g., body motion).

5.2. Instruction of newcomer sensors

In case the replaced sensor is different from the predecessor one, or a new sensor is introduced in the network, a complete redefinition and retraining of the AR system is required. Likewise, if recognition capabilities want to be given to a sensor originally not devised for behavior-awareness tasks, a new model must be built according to their particular characteristics. Following a classic learning process is in both cases quite costly, since it normally requires to collect new experimental data. The collection of new data implies to record the behavior of a person or set of people while performing the activities of interest for the new sensor setup. Apart from being a very long and tedious process, sensor setups may vary from person to person, or even from one context to another, thus this approach proves impractical for real-world settings.

The training of newcomer sensors should be performed without the involvement of a system designer, which otherwise would limit the approach to predefined sensor setups and deployments. This must also happen without user involvement. To fulfill these requirements, the most reasonable approach is to use the actual knowledge of the existing AR system to instruct the new sensors on the activity-awareness tasks. This is accomplished in a process in which the original AR system or "teacher" transfers its knowledge to the newcomer untrained sensor or "learner". Also known as inductive transfer or transfer learning [91], this corresponds to a research problem in machine learning that focuses on translating the knowledge available to solve a problem in one domain to a different but related domain.

Transfer learning has been rarely approached in AR, although a few recent works have contributed to this respect (see Section 2.5). A naive approach to pursue transfer learning consists in directly utilizing the activity model defined for a particular sensor on the data measured through the newcomer sensor. This may be performed in a fast way, and no specific training of the newcomer system is required. However, this approach suffers from important limitations. First, even assuming that both teacher and learner sensors are of identical characteristics and modality, if the newcomer sensor is devised to be placed in a body part

distant from the teacher default placement (e.g., wrist sensor to be the teacher and upper arm sensor to be the learner), the translated model may be expected not to suit for the purpose of recognition. Moreover, if teacher and learner build on sensors of the same modality but different technical characteristics, it could also be expected that this approach fails to provide an accurate transfer of activity-awareness. More importantly, the predefined AR system cannot be in principle used for the newcomer sensor if this is of a different modality to the teacher sensor (e.g., a system devised to operate on acceleration data is used to work on the signals measured through a gyroscope).

An alternative consists in transferring the capacity of recognizing activities of an existing sensor to another untrained node through an online supervised learning, where the supervision (i.e., the labels) is provided by the former. To that end, the activity class identified by the teacher sensor is employed to tag or label the data measured at each time by the learner sensor. Through this model, newcomer sensors could be learned even when these differ in characteristics or modality with respect to the teacher sensors. However, this approach also presents important drawbacks that complicates its use in realistic settings. Firstly, only the data of those activities performed by the user could be tagged. This is very limiting, since some activities are more habitually performed than others during daily living conditions. For example, in long-term behavioral monitoring only a few activities occur often (e.g., sleeping, sitting), whereas most activities happen rather occasionally (e.g., open a window). In this regard, the transfer learning may take long time spans to address the complete activity set, and at worst be incomplete. This also leads to an activity imbalancement that may bias the system capabilities during learning. The learning process is further coupled to the performance of the teacher system. Errors made during the recognition of actual activities are propagated to the learner system through the labeling process. This determines that the performance of the newcomer will be in general less or equal to the one achieved by the teacher system. The situation becomes even worse when these errors are not only associated to the normal functioning of the recognizer but also to sensor anomalies such as faults or displacements. In that case, the newcomer system may learn the anomalous behavior of the teacher system, thus leading to a malfunctioning of the learner system. At worst, the teacher system could not provide tags as it is incapable of activity recognition (e.g., sensor out of battery). Under these circumstances, the learning process may be interrupted or

not performed at all.

Taking into account the shortcomings of the previous models, it seems clear that transfer learning for real-world conditions requires to be performed in an autonomous and rapid fashion and apply to heterogeneous sensor configurations. Through the use of transfer learning concepts, this work aims at making it possible to apply the activity models of a predefined sensor system or source to the data from a newly discovered sensor system or target that lacks of such models. It is assumed that both systems coexist for a short time and that there is a (unknown, to be determined) function that allows for mapping between the signals of source and target sensor systems. In the following, the models proposed to identify these relations as well as to transfer the AR models are presented.

5.3. Multimodal transfer methods

The idea of transfer learning is here meant as the means to translate a source system capable of activity recognition (teacher) to a target system which lacks this ability (learner). Two transfer learning methods are proposed depending on whether the source or teacher AR system is defined through activity templates (signal patterns) or activity models (feature/classification models).

The proposed transfer methods work in two steps. First, a system identification technique (Section 5.3.1) finds a function that maps the signals of the source sensor to the signals of the target sensor. This process can be carried out on signals corresponding to different domains or modalities (cross transfer) or between data corresponding to the same domain (identical transfer). Based on this mapping, the AR system is then translated. For the AR system based on templates (Section 5.3.2) the translation process consists in transforming the activity templates from the source to the target domain. These new templates may be used for recognition or to build an activity model for the target sensor. For the recognition system based on activity models (Section 5.3.3) the translation process consists in literally conveying or copying the source activity model to the target system. To operate on this activity model the target system needs to map its signals to the source domain in which the activity model was originally defined.

5.3.1. System identification

The complexity of the signal mapping transfer starts from the physics of the domain transformation. In some cases the underlying relationship is well-defined and known (e.g., position to acceleration) but for other cases may not be clearly identifiable (e.g., position to magnetic field). In addition, the setup constraints associated to each particular context and subject reduce the generalization capabilities of those models which are problem-specific defined. In many cases such models will be overly complex and almost impossible to obtain in reasonable time due to the complex nature of many systems and processes. Furthermore, the complexity of this engineering design increases dramatically when the number of sensors also grows, thus constituting a non scalable approach.

The transformation or mapping of data from one domain to a similar or different domain is normally approached in base of the a priori knowledge of the underlying relationship among domains. White-box (WB) models, also known as user-defined models, are usually selected when all the necessary information about the problem domain is available. Hence, WB designs require to know for instance the placement and nature of the sensors, the features delivered by each sensor node (data range, units) or the number and type of sensors (modalities) in order to define an appropriate mapping transfer function. This task is extremely time and resource-consuming because a design team is required to analyze and implement the models for each combination of sensors and domains. Moreover, this approach does not fulfill the desired autonomous characteristic of the system, where no prior knowledge of the problem domain and sensor ecosystem must be assumed.

Black-box (BB) models are encountered to be more appropriate for this problem as they are data-driven models where only the inputs (source) and outputs (target) are needed, with limited additional knowledge about the internals of the model [113]. BB modeling is often employed when assumptions on the nature of the underlying system are hard to make, when the complexity of the underlying relation is extremely high, or to avoid designer's bias, which fits in well with the characteristics of this problem. The choice of the mathematical functions within BB models is normally made depending on whether the system to model is linear or non-linear, and whether it is time-dependent or time-independent. A system can also have single or multiple inputs and/or outputs or explicitly incorporate external exogenous disturbances such as noise. In summary, the most interesting charac-

teristics of the use of BB models instead of a classical WB approach are:

- *Generalization capabilities*: it can be in principle used for whichever kind of mapping, even combining different type of domain sources or using abstract magnitudes, such as proximity or object-interaction sensors (e.g., switch buttons, RFID).
- *Scalability*: in general the model may be applied with no much effort to a larger set of sensors or signals, i.e., inputs, of different nature.
- *Robustness to information loss*: the statistical capabilities of some of its regression and prediction models allow for learning even when data loss events are present.
- *Design complexity*: there is no need to explicitly discover the underlying relation that links the domain transformation; no extra information about the sensor deployment is a priori required; reduced design time, normally automatic and autonomously performed.
- *User abstraction*: no specific user intervention is required, thus keeping aside burdensome data collection procedures for training of the systems.

The field of system identification provides techniques to build models[1] of dynamical systems from data [114, 115, 116, 117, 118]. Examples of these techniques are state-space models [119], which use state variables to describe a system by a set of first-order differential or difference equations. Artificial neural networks (ANNs) offer rich sets of transfer functions such as linear or non-linear time-invariant functions (e.g., with multi-layer perceptrons) and time-variant functions (e.g., with time-delay neural networks or recurrent neural networks), but at the expense of large training data and slow training process [120]. Autoregressive-moving-average (ARMA) models represent a family of

[1]Note the distinction between models used for AR (here "activity models") and models resulting from system identification ("system identification model" or "mapping"). The latter is meant here. It is a mathematical description of the relation between quantities of a physical system, such as the readings delivered by multiple sensors.

tools to predict future values of a time series [113]. They are commonly used in signal processing, speech processing and automatic control. They offer more constrained transfer functions and have the advantage of using less training data and relatively simple learning rules (least square regression). Exogenous or independent inputs may also be modeled through a generalization of the ARMA model (autoregressive-moving-average with exogenous inputs (ARMAX)). ARMA and ARMAX models are commonly used for time-series forecasting, in applications as diverse as electric load [121], water resources [122], or mechanical structures [123].

The system identification model should allow for transformations between the habitual sensing modalities that are used for AR, most of them identified as of a linear nature. Some typical static transformations include scaling (sensors with different sensitivity or units - absolute/relative -) and affine transformations, offset (different zero value), non-linearity (compression of the dynamic range), translation or rotation (e.g., when an acceleration sensor is displaced - translated and/or rotated -). Dynamic transformations may include multiple differentiation or integration operations (e.g., between position or angle and linear or angular velocity), or hysteresis. Besides, most sensors utilized in AR problems measure various axes at once (e.g., tri-axial accelerometers, positioning systems, etc.) which are not independent, so the system identification model has to be of the multiple-input-multiple-output (MIMO) kind.

Accordingly, this work proposes the use of a parametric linear model, which surpasses non-linear approaches in terms of [117]: interpretation (i.e., represents and extracts the properties and knowledge of the underlying relationship); generalization (i.e., captures the true dynamics and predicts accurately the output for unseen new data); robustness to overfitting and noise rejection; speed; and amount of data required for the training and complexity (i.e., training time, computational resources required, etc.). Here a linear MIMO mapping is specifically used for system identification [124]. Such mappings can be directly learned from data. This permits to learn mappings in a wide range of sensing environments without designer involvement or bias. In the following, the mathematical description of the employed models is provided.

Let us define $\mathbf{x_S}(t)$ as an n_S-by-1 vector of sensor data from the source domain S at time t and $\mathbf{x_T}(t)$ as an n_T-by-1 vector of data of the sensors of the target domain T. A mapping relating the sensor signals in

different domains is first identified. This may be from source to target signals, or target to source signals, whichever can be best identified. Let us denote with $\Psi_{S\to T}$ the function that maps the source to the target signal: $\Psi_{S\to T} : \mathbf{x_S}(t) \to \hat{\mathbf{x_T}}(t) \approx \mathbf{x_T}(t)$. $\Psi_{T\to S}$ defines as the function that maps the target to the source signal: $\Psi_{T\to S} : \mathbf{x_T}(t) \to \hat{\mathbf{x_S}}(t) \approx \mathbf{x_S}(t)$. The $\hat{}$ symbol is used to indicate that the signal is predicted in a given domain from the known signal of another domain.

A linear MIMO mapping is defined as follows:

$$\mathbf{x_T}(t) = \mathbf{B}(l)\mathbf{x_S}(t) \tag{5.1}$$

where $\mathbf{B}(l)$ is a n_T-by-n_S polynomial matrix in the delay operator l^{-1} (i.e., each entry of the matrix is a polynomial in l^{-1}). The operator l^{-k} introduces a delay of k samples in the signal to which it is applied: $l^{-k}x(t) = x(t-k)$. The source and target sensor signals are the inputs and outputs of the model. The matrix $\mathbf{B}(l)$ contains elements $b_{ik}(l)$,

$$B(l) = \begin{pmatrix} b_{11}(l) & b_{12}(l) & \cdots & b_{1n_S}(l) \\ b_{21}(l) & b_{22}(l) & \cdots & b_{2n_S}(l) \\ \vdots & \vdots & \ddots & \vdots \\ b_{n_T1}(l) & b_{n_T2}(l) & \cdots & b_{n_Tn_S}(l) \end{pmatrix} \tag{5.2}$$

of the form:

$$b_{ik}(l) = b_{ik}^{(0)} l^{-s_{ik}} + b_{ik}^{(1)} l^{-s_{ik}-1} + \ldots + b_{ik}^{(q)} l^{-s_{ik}-q} \tag{5.3}$$

where q is the number of past input samples that are used for the computation of the current output sample and s_{ik} are the static delays from the k-th input to the i-th output. Otherwise, $b_{ik}(l)$ represents the transfer function in the Z-transform domain from the k-th input (k-th channel of the source system) to the i-th output (i-th channel of the target system). In this way, $\mathbf{B}(l)$ accounts for the contributions of all inputs to calculate the outputs. For the identification of the $(q+1) \times n_T \times n_S$ coefficients of the polynomials and the $n_T \times n_S$ static delays, a least squares approach is followed. QR factorization solves the overdetermined set of linear equations that constitutes the least-squares estimation problem. Internal loop feedback is here not considered for the sake of simplicity, thus the transfer function is rather devised as a forward combination of the inputs, i.e., a linear combination of the tapped delay inputs.

The linear MIMO mapping allows for combinations of subsets of the transformations mentioned above:

- Scaling. This is obtained by setting $b_{ik}^{(0)}$ to the scaling factor and $b_{ik}^{(s)}$ to zero $\forall\, s > 0, \forall\, i = k$. Furthermore, all the coefficients $b_{ik}^{(s)}, i \neq k$ of the off-diagonal polynomials will be zero, yielding a diagonal matrix.

- Rotation. This is obtained by setting $b_{ik}^{(0)}$ to the corresponding element at position ik in the rotation matrix and by setting $b_{ik}^{(s)}$ to zero $\forall\, s > 0$.

- Differentiation of order h. This is obtained by setting $b_{ik}^{(s)}, \forall s \leq h$, $\forall\, i = k$ to the corresponding coefficients of the transfer function of the derivative. All the other coefficients are set to zero.

5.3.2. Transfer of activity templates

Given a source recognition system defined through activity templates[2], the objective here is to transfer this system recognition capabilities to a new target system. To that end, the source system activity templates need to be translated to the target system domain, where the latter may use them for AR.

The complete architecture of the method is depicted in Figure 5.1. It starts from a source operational AR system (i.e., trained) that recognizes activities (act) from the data of a sensor (encircled S). The recognition system devised for the source domain also stores the activity templates $\mathcal{T}_S$. $\mathcal{T}_S$ consists of raw sensor signals $\mathbf{x_S}(t)$ and the corresponding class labels. First, a mapping function $\Psi_{S\rightarrow T}$ between source and target sensor signals is obtained through system identification (Section 5.3.1). Then, $\Psi_{S\rightarrow T}$ is used to translate the templates $\mathcal{T}_S$ into templates $\mathcal{T}_T$ containing the predicted sensor signals $\hat{\mathbf{x_T}}(t)$ in the target domain, and the corresponding class labels. The target system trains its AR system based on $\mathcal{T}_T$ (e.g., running a feature extraction and selection process and training a classifier based on $\mathcal{T}_T$). At this point the target system is ready to operate on the data of domain T.

[2]Activity templates are signal patterns that represent certain activities or gestures. These patterns may be directly used for signal recognition or pattern matching (dynamic time warping [125], shapelets [126], etc.) or for building more sophisticated activity models through feature extraction and classification procedures.

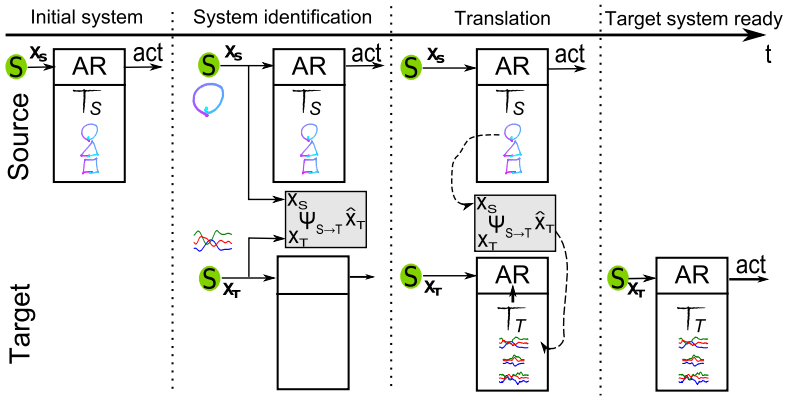


Figure 5.1: *Architecture for the transfer of activity templates. From left to right, initially a fully operational AR system defined through activity templates is identified. Second, a mapping function between source and target domains is discovered through system identification. Thereafter, the activity templates are translated from source to target domain, thus allowing the target system to use the translated templates to build its own AR system. Finally, the target system is ready to operate. Note: the depicted signals may for example represent position (source domain) and acceleration (target domain).*

5.3.3. Transfer of activity models

In this case, the recognition system devised for the source domain relies on activity models $\mathcal{M}_S$ (i.e., the parameters of the recognition system, including the selected set of features, the trained classifiers, etc.). The idea is that the target uses the same activity models ($\mathcal{M}_T = \mathcal{M}_S$). Therefore, the translation process basically consists in copying $\mathcal{M}_S$ to the target model. However, to be capable of using $\mathcal{M}_T$ the target system requires their signals to be translated into the source domain. To do so, the target system uses $\Psi_{T\to S}$ to translate the sensor signals of domain T ($\mathbf{x_T}(t)$) into domain S ($\hat{\mathbf{x_S}}(t)$) prior to applying the AR model. Here again, $\Psi_{T\to S}$ is obtained through system identification (Section 5.3.1). The complete activity model transfer architecture is shown in Figure 5.2.

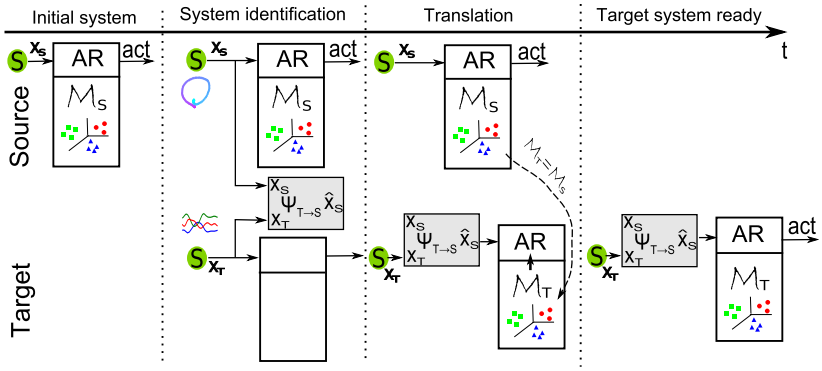


Figure 5.2: *Architecture for the transfer of activity models. From left to right, initially a fully operational AR system defined through activity models is identified. Then, a mapping function between target and source domains is discovered through system identification. Thereafter, the source activity models are translated to the target domain so both use the same activity models. These activity models also define the target AR system. Finally, the target system continuously translate their signals into the source domain to operate the recognition system. Note: the depicted signals may for example represent position (source domain) and acceleration (target domain).*

5.4. Evaluation of multimodal transfer

To evaluate the capabilities of the proposed transfer methods a multimodal setting is particularly considered. This consists of a gesture-based HCI scenario in which AR capabilities are transferred between sensors of the same modality (between body-worn and body-worn sensors, i.e., identical transfer) and different modality (between body-worn and ambient sensors, i.e., cross transfer). The body-worn sensors are five IMUs of which only the acceleration is used. The ambient sensor is a consumer vision-based skeleton tracking system (Microsoft Kinect) that provides the position of the body joints. The fact that the Kinect sensor normally builds on activity templates, and AR systems based on IMU sensors operate on either activity templates or activity models, makes of this a fairly setup to evaluate both transfer methods.

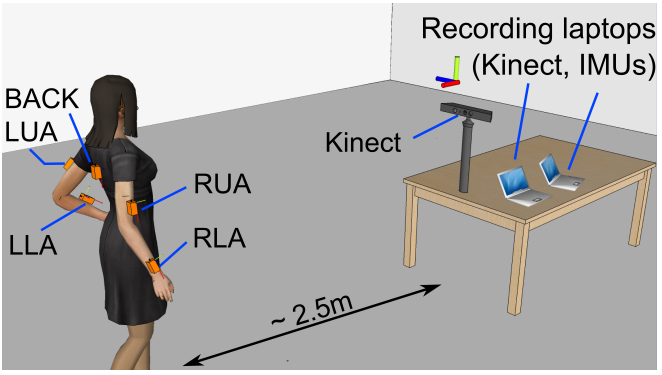
5.4.1. Experimental setup and dataset

The test bench for this work is a gesture recognition setup (Figure 5.3) with five body-worn IMUs and a consumer vision-based skeleton tracking system (Microsoft Kinect). These sensors are commonly deployed for AR. IMUs are widely available on smartphones and can be highly miniaturized, e.g., for integration in garments [127]. Kinect allows activity-aware gaming on the Xbox console. Apart from the accuracy and precision of the Kinect model to obtain a reliable description of the body model, the idea of using this system is supported by the low cost, setting simplicity (plug-and-play) and increasingly developer support which constitute it as the hottest platform in the industry for building new inventions and the fastest selling electronics device ever[3]. Obviously other approaches based on general purpose cameras can be similarly considered, but the user accessibility and research potential for Kinect is nowadays incomparable. In AR the Kinect sensor has been used for example for the recognition of activities of daily living [128] and gait analysis [129].

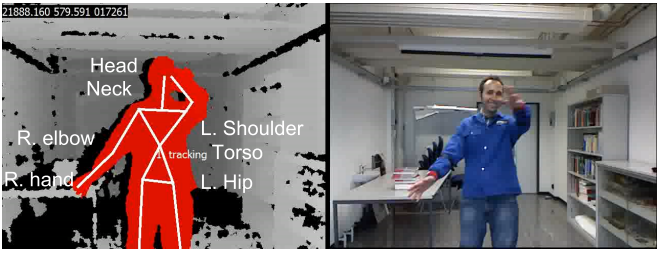
Kinect contains an 8-bit 640×480 RGB camera, an infrared (IR) LED projecting structured light and an IR camera. It computes on-the-fly an 11-bit 640x480 depth map in a range of 0.7-6m from the reflected IR light. The drivers fit a 15-joint skeleton on the depth map (proprietary algorithm similar to [130]) in real time and deliver 3D joint coordinates in millimeters measured from the Kinect center. Tracking is specified in a range of 1.2-3.5 meters [131]. Kinect is interfaced over USB to a PC. The RGB and depth map videos and the joint coordinates are recorded at 30Hz by using [132]. The 3D coordinates of the joints and center of mass are provided in millimeters from a Kinect-centered coordinates system together with an estimate of the position accuracy between 0 and 1. This permits to filter unreliable data.

Five IMUs (XSens [133]) wired to a PC sense the upper body orientation. These IMUs contain gyroscope, magnetometer and acceleration sensors combined with a Kalman filter to yield the device orientation in a world coordinate system in real-time. Here [112] is used to acquire the raw sensor data and the device orientation at 30Hz. For the sake of this evaluation only the 3D (tri-axial) acceleration measured by the IMUs is used.

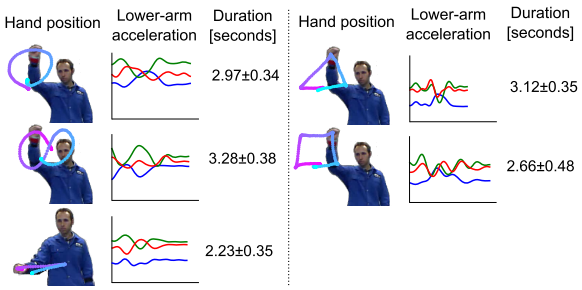
[3]Kinect sold 10 millions units between its release on November 4th, 2010 and March 2011, earning it the Guinness World Record of the "Fastest selling consumer electronic device". Its low cost (150$) puts it in the reach of many households.



(a) Experimental Setup



(b) Kinect depth map and skeleton (left) and RGB (right)



(c) HCI gestures (position, acceleration)

Figure 5.3: *Kinect and IMU experimental setup. (**a**) IMUs and a Kinect capture the user's movements. (**b**) The Kinect sensor delivers a depth map, a color image and a 15-joint skeleton of the user. (**c**) The right hand position and limb acceleration are synchronously recorded for five gesture kinds.*

The Kinect and IMU data are independently recorded and resampled offline to the regular Kinect sampling rate to obtain a synchronized dataset[4] comprising acceleration, position and labels. A single subject performs five kinds of geometric gestures (circle, infinity, slider, triangle, square) with the right hand in alternation 48 times (Figure 5.3(c)). These gestures were selected because similar ones were demonstrated to be recognizable by wearable sensors [78, 134, 76] or Kinect [131]. They involve the lower and upper arm, which permits to assess the approach for the transfer of AR systems between limbs. They are also diverse enough, which allows us to study if there are preferential movements leading to a faster identification of the mapping between the two systems. The average±standard deviation of the duration of each gesture classes are 2.97(89)±0.34(10), 3.28(98)±0.38(11), 2.23(66)±0.35(10), 3.12(93)±0.35(10) and 2.66(79)±0.48(14) seconds(samples). A five minutes long "idle" dataset, where the user performs infrequent low-amplitude arm movements and moves around, without any specific task is also recorded. The subject is between 2-3 meters from the Kinect sensor facing it within ±30° to avoid occlusions (Figure 5.3(a)). Annotation was performed on-the-fly and corrected later using video footage of the Kinect (Figure 5.3(b)). Hand-claps at the start and end of the recording are also used for offline synchronization.

No constraints were placed on the way the gestures are executed with the exception that the subject should try his best to execute them in a similar manner. However, it was observed that gestures were executed somewhat faster later in the recording, and that the user position shifted away from the center of the camera field of view, until the user consciously moved back to the center. The left arm does not experience significant movement, thus the information monitored is not considered for this particular experiment. Moreover, since the ultimate goal would be to perform the transfer in a short time, for example, just by performing a reduced subset of informative gestures (ideally just one single gesture) to learn the mapping, this scenario is considered the most adequate for that purpose. One subject is just involved since the mapping should be learned with independence of who performs the movements.

The capacity of the transfer methods is assessed through two metrics. First, the system identification performance is evaluated by assessing the quality or fit of the signal $\hat{x_T}$, obtained by mapping x_S to the target domain with the MIMO model, compared to the measured signal x_T. Second, it is assessed the accuracy with which the system can

[4]The dataset is accessible at http://www.ugr.es/~oresti/datasets.

classify the gestures after transfer to the target domain T, compared to the accuracy in the source domain S which is used as a baseline.

The fit[5] between the measured on-body acceleration $\mathbf{x_T} = \mathbf{x_I}$ and the predicted acceleration $\hat{\mathbf{x_T}} = \hat{\mathbf{x_I}}$, obtained by mapping the source signals $\mathbf{x_K}$ to the target domain is calculated for each channel i:

$$BestFit_i = 1 - \frac{\left(\sum_{t=1}^{N}\left(x_T^{(i)}(t) - \hat{x_T}^{(i)}(t)\right)^2\right)^{\frac{1}{2}}}{\left(\sum_{t=1}^{N}\left(x_T^{(i)}(t) - \bar{x_T}^{(i)}\right)^2\right)^{\frac{1}{2}}} \tag{5.4}$$

with N the number of signal samples, and $\bar{x_T}^{(i)}$ the mean over time of $x_T^{(i)}(t)$. The $BestFit_i$ is then averaged on all channels, resulting in a unified *BestFit* value. A *BestFit* of 1 indicates a perfect fit. Values above zero qualitatively indicate a good fit (Figure 5.4). As *BestFit* tends to $-\infty$, the prediction differs more and more from the target.

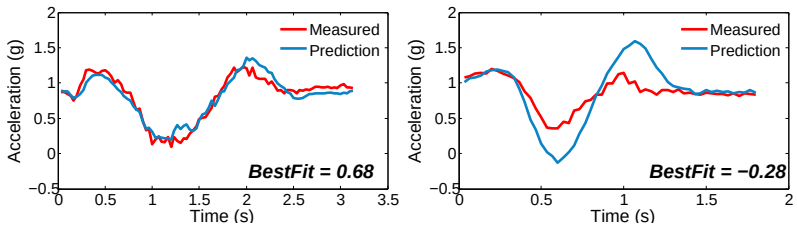


Figure 5.4: *Comparison between the actual acceleration measured at the lower arm and the predicted after mapping from the hand position sensed by Kinect, for a circle (left) and a slider (right). Visually, a good match between predicted and measured signals is obtained for BestFit values above 0.*

Three kinds of MIMO mappings are evaluated in this work:

- *Problem-domain mapping (PDM).* This is a generic mapping learned on instances of all classes in equal proportions.

[5]Mean square error (MSE) and root mean square error (RMSE) are commonly used to measure the fit in a regression problem. These metrics however are highly affected by scale and offset of the signals. This may be partially addressed by the normalized variants (NMSE and NRMSE) and mean subtraction, which is essentially used in the *BestFit* definition. Other similarity measures such as correlation or mutual information could be also used.

- *Gesture-specific mapping (GSM).* This is a mapping learned on instances of a single class. It is used to analyze whether specific movements are more suited to identify the system dynamics.

- *Unrelated-domain mapping (UDM).* This is a mapping learned from a sequence of samples from the idle dataset. It is used to assess mapping generalization across scenarios.

The models are trained with a minimum of data corresponding to roughly the duration of the longest gesture, which is set to 100 samples. Thus GSM and UDM are learned on 100 samples and PDM on 500 (as it requires data of each of the 5 gestures). The MIMO mappings are learned on a subset of the dataset and evaluated on the rest. The learning subset is obtained from a particular instance (GSM), by aggregating multiple activity instances (PDM), or obtained from the idle dataset (UDM). For this evaluation the selection is randomly repeated 20 times in an outer cross-validation process.

To evaluate[6] the accuracy of the transfer three non-overlapping parts of the dataset are used to learn the MIMO mapping, to train the source recognition system, and to test the translated target recognition system. The classifier training and testing sets are defined by an instance-based random-seed 10-fold inner cross-validation process, repeated 100 times. The data used to learn the MIMO mapping is selected as indicated previously in an outer cross-validation process. Source and target baseline classification accuracies are assessed by training and testing on data from the same domain. For the transfer of activity templates, the translation is evaluated by training the recognition system on the predicted acceleration and testing it on the measured acceleration. For the transfer of activity models, the translation is evaluated by training the recognition system on the measured acceleration and testing it on the predicted acceleration.

Two feature sets are used. Each instance is subdivided into 4 subwindows that capture the temporal dynamics and features are computed on them. Typical features used in AR systems are employed. Concretely, FS1 corresponds to the mean of each axis (12 features) while FS2 is the maximum and minimum of each axis (24 features).

[6] All the processing is performed in Matlab R2011b. For the system identification some of the functions provided in the *System Identification* toolbox have been used, while many others have been specifically defined for this purpose. Signal processing, featuring and classification is performed through the use of models similar to the used in previous work of this thesis.

The accuracy for segmented gestures recognition with a KNN classifier is reported. A model similar to the one proposed in [103] is here used. The k-value for the KNN model is set to three for the sake of simplicity and given the good results shown for previous work of this thesis. In fact, KNN models have been proven to perform well in gesture recognition for both Kinect [135] and IMUs [136, 28].

5.4.2. Transfer between IMU and IMU

Here the transfer of the recognition capabilities of an existing AR system operating on an IMU to a new untrained system devised to operate on a different IMU is investigated. For the sake of simplicity, the source recognition system will be identified in advance as I_S and the target as I_T. The translation between IMU and IMU relies on the identification of $\Psi_{I_S \to I_T}$ or $\Psi_{I_T \to I_S}$ which correspond to a 3-input (3D acceleration) 3-output (3D acceleration) MIMO mapping with 10 tap delays (q = 10, n_S=3, n_T=3, implies 108 parameters to learn, which refers to the $(q+1) \times n_T \times n_S$ coefficients of the polynomials and $n_T \times n_S$ static delays, as described in Section 5.3.1). The tap delay is set to this value to procure that the MIMO model captures the dynamics of the transformation. Although the overfitting of the models potentially increase with the number of parameters, this is here avoided through the cross-validation procedure.

In Section 5.3 two methods that allow us to transfer AR systems through the exchange of either activity templates ($\mathcal{T}$) or activity models ($\mathcal{M}$) were presented. Now, the specialization of these techniques to the IMU and IMU test case is described. AR systems based on IMU sensors may operate on both activity templates $\mathcal{T}$ and activity models $\mathcal{M}$. Therefore, the two types of transfers are evaluated.

First of all, the signal mapping from the source IMU (acceleration) to the target IMU (acceleration) is identified. The **System Identification** process characterizes through:

- A MIMO mapping $\Psi_{I_S \to I_T}$ from the 3D acceleration of the source IMU to the 3D acceleration of the target IMU is learned for the transfer of activity templates.
- The reverse MIMO mapping, $\Psi_{I_T \to I_S}$, is needed for the transfer of activity models.

For the **Transfer of Activity Templates:**

- The source domain recognition system works on the 3D acceleration measured by the source IMU. It also stores the activity templates $\mathcal{T}_S = \mathcal{T}_{I_S}$ that are the 3D acceleration patterns for each gesture.
- $\mathbf{x_S} = \mathbf{x_{I_S}}$ is the 3D acceleration measured on the body by the source IMU.
- $\mathbf{x_T} = \mathbf{x_{I_T}}$ is the 3D acceleration measured on the body by the target IMU.
- $\hat{\mathbf{x_T}} = \hat{\mathbf{x_{I_T}}} = \Psi_{I_S \to I_T}(\mathbf{x_{I_S}})$ is the acceleration predicted on the body from the known source acceleration.
- After template translation, $\mathcal{T}_T = \mathcal{T}_{I_T}$ are the predicted 3D on-body acceleration patterns for each gesture and the corresponding class labels.
- The target recognition system is automatically trained at runtime on the templates $\mathcal{T}_T$, and finally operates on the acceleration sensed by the target IMUs.

For the **Transfer of Activity Models:**

- The source domain recognition system works on the 3D acceleration sensed by the source IMU. It uses models $\mathcal{M}_S = \mathcal{M}_{I_S}$ for the recognition.
- $\mathbf{x_S} = \mathbf{x_{I_S}}$ is the 3D acceleration measured on the body by the source IMU.
- $\mathbf{x_T} = \mathbf{x_{I_T}}$ is the 3D acceleration measured on the body by the target IMU.
- $\hat{\mathbf{x_S}} = \hat{\mathbf{x_{I_S}}} = \Psi_{I_T \to I_S}(\mathbf{x_{I_T}})$ is the acceleration predicted on the body from the unknown source acceleration.
- After translation, the 3D acceleration of the target IMU are mapped to "look like" the acceleration measured on the source IMU. In this way, the recognition models devised for the source IMU are used as is by the target system ($\mathcal{M}_T = \mathcal{M}_S$) that now operates on the target IMU data.

The transfer among all possible pair combinations of the IMU sensors placed on the user right lower arm (RLA), right upper arm (RUA) and back (BACK) are considered for evaluation. This leads to six cases of transfer of AR from a source to target domain, concretely from RLA to RUA, RLA to BACK, RUA to RLA, RUA to BACK, BACK to RLA and BACK to RUA. For all these combinations both transfer of activity templates and activity models are analyzed. These scenarios are devised to help investigate the potential of the transfer methods for sensor systems of same modality (acceleration) and of diverse characteristics (centered on close-by or unrelated body parts).

System identification performance

The *BestFit*[7] computed from the evaluation of all possible pair combinations of mappings between RLA, RUA and BACK is depicted in Figure 5.5. For example, the *BestFit* for the RLA to RUA mapping is computed between the acceleration measured at the lower arm and the acceleration predicted from the upper arm. From the results, the best fit tends to be obtained with PDM (Figure 5.5(a)). For this case, almost all *BestFit* values are above 0, which was shown in Section **??** to correspond to a good fit. Moreover, for some mappings the *BestFit* distributions are close to 1, which represents an almost perfect mapping. This may be expected, as the mappings are learned on the dynamics of all gestures. The results are also very promising for the GSM model. Once again, *BestFit* values superior to 0 are generally obtained but for triangle and square gestures for some sensor combinations. A good fit is obtained for the rest of gestures, thus one of them could be in principle enough to learn a mapping model. Finally, the UDM model provides the worst fit performances (Figure 5.5(c)). Although high *BestFit* values are obtained for some mappings, these are in many cases below 0. The large dispersion in the results demonstrate that some of the movements performed while idling may be used for learning a mapping, however, many others fail to provide valuable information for capturing the dynamics of the physical system. This is consistent with the characteristics of the idle dataset, since it has only rare occurrences of large amplitude limb movements. However, this does not exclude that a dataset from a domain not comprising the activities to

[7]The interest of these results is not focused on the analysis of particular *BestFit* values but on the comparison of the statistical distribution for each case.

recognize, but containing richer limb movements, may not be used to learn an adequate mapping model.

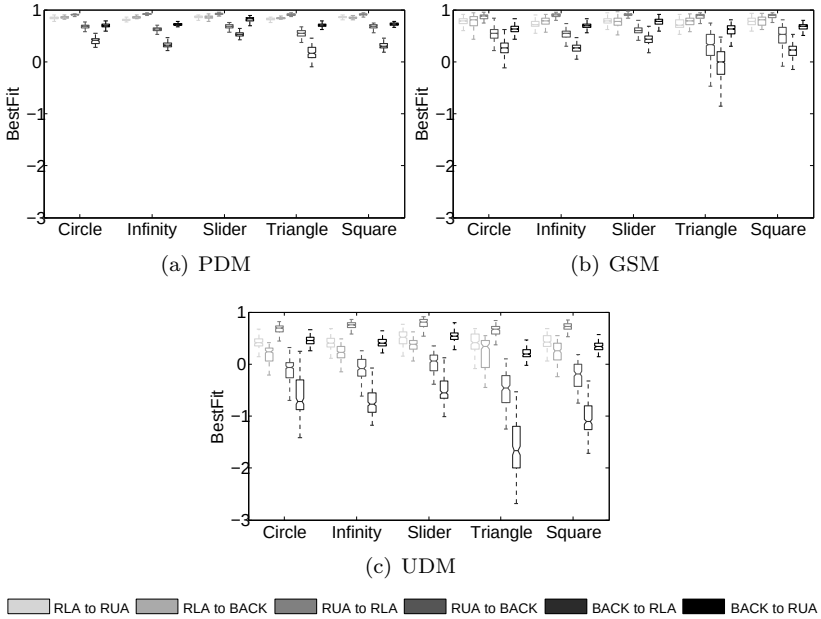
Best fits are obtained between close-by limbs. Concretely, mappings from the lower arm to the upper arm and vice versa obtain the highest *BestFit* values for all types of mappings. Likewise, mappings between upper arm and back prove to be good-enough. The fit worsens as the mapping model is computed between less related body regions (back to lower arm). This seems to be quite reasonable since lower arm and upper arm movements, and upper arm and back movements are more related to each other than between lower arm and back.

Transfer accuracy

Source and target baseline classification accuracies are assessed by training and testing on data from the same domain. The IMU to IMU transfer of activity templates is evaluated by training the recognition system on the predicted acceleration $\hat{\mathbf{x}_{\mathbf{I_T}}} = \Psi_{I_S \rightarrow I_T}(\mathbf{x}_{\mathbf{I_S}})$ and testing it on the measured acceleration $\mathbf{x}_{\mathbf{I_T}}$. The IMU to IMU transfer of activity models is evaluated by training the recognition system on the measured acceleration $\mathbf{x}_{\mathbf{I_S}}$ and testing on the predicted acceleration $\hat{\mathbf{x}_{\mathbf{I_S}}} = \Psi_{I_T \rightarrow I_S}(\mathbf{x}_{\mathbf{I_T}})$.

Classification accuracy baselines in the source (BS) and target domain (BT), and those after transfer to the target domain are presented in Figure 5.6 for FS2 (this set is used because it is more sensitive to inaccurate signal mapping). The baselines represent the accuracy obtained by a recognition system that is trained and tested on the same sensor (no transfer). The GSM mapping is learned on the "circle" gesture, which was identified as one of the best gestures to learn a mapping model. The baselines indicate that the gestures can be classified with an accuracy of 98% or more with the lower-arm acceleration and the upper-arm acceleration. The high accuracy obtained with the back acceleration (baseline of about 89%) indicates that torso movements are correlated with the execution of the gestures. This is a particular characteristic of this scenario, that likely does not generalize to other scenarios. The results after transfer must be assessed according to the performance drop from the baselines. The performance drop from BS indicates how much worse the system becomes after transfer. The drop from BT indicates how much better would be a system devised specifically for the target domain.

The transfer between close-by limbs proves to be the most efficient.

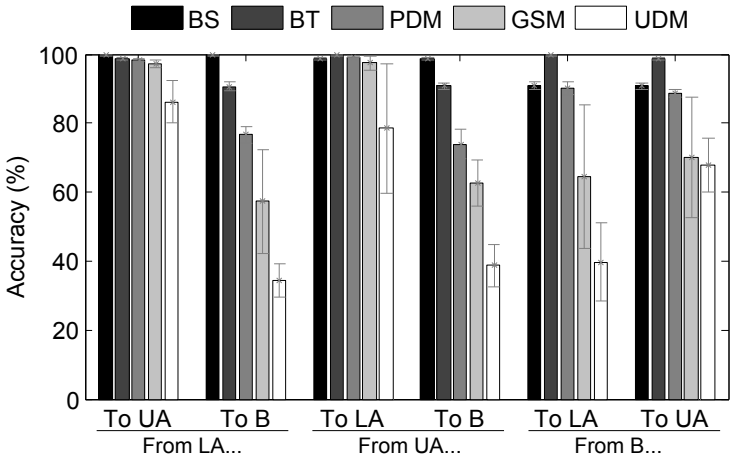


(a) PDM (b) GSM

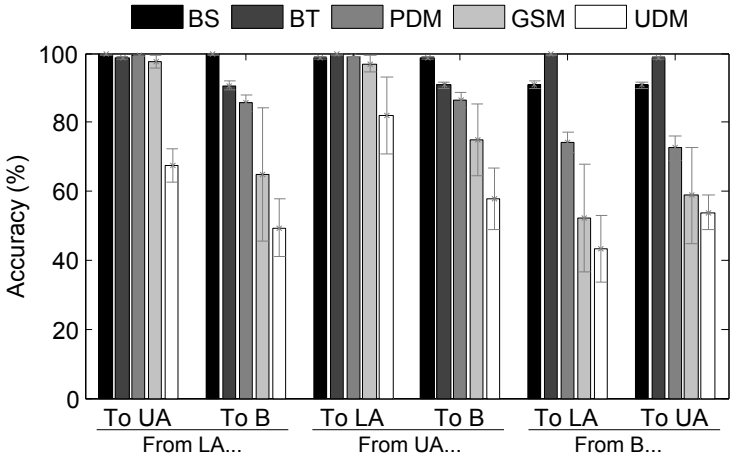
(c) UDM

Figure 5.5: *Box plot of the BestFit distributions for all possible pair combinations of mappings between RLA, RUA and BACK. The box plot represents the statistic distribution of the sample set (the central mark is the median, the edges of the box are the 25th and 75th percentiles, and the whiskers the most extreme results not considered outliers). BestFit for RLA to RUA, RLA to BACK, RUA to RLA, RUA to BACK, BACK to RLA and BACK to RUA mappings are respectively represented by each box within each gesture group. (**a**) The mapping is trained on all gestures and the fit computed on the indicated gestures. (**b**) The mapping is trained on the indicated gesture and the fit computed on all of them. (**c**) The mapping is trained on data from another domain, and the fit is computed on the indicated gestures.*

Accordingly, best results are obtained for the translation between upper arm and lower arm. For both PDM and GSM models the transfer from the lower arm to the upper arm is practically similar to when performed the other way around. Moreover, almost no drop is observed after transfer for these combinations. In fact, the performance obtained when using the PDM model is similar to baseline, and reduces less than



(a) Transfer of activity templates



(b) Transfer of activity models

Figure 5.6: *Classification accuracy (average - bar - and standard deviation - whiskers -) for the translation between two IMU systems with FS2. Transfer* `from` *a source system operating on* ***(a)*** *activity templates or* ***(b)*** *activity models* `to` *an untrained new system. Source and target systems are respectively identified through the X-axis. BS and BT indicate the baseline accuracies obtained with a system trained and tested on the source and target domain.*

3% at worst for the GSM model. This implies that executing a single "circle" is sufficient to identify a reliable mapping model. Transfers between distant body parts provide in general the worst results. The transfer between the lower arm and the back acceleration shows a variable drop from baseline depending on the mapping (10% to 65% for the activity templates and 5% to 55% for the transfer of activity models). This is observed for both directions of the transfer. Better results are obtained when translating recognition capabilities between upper arm and back systems. This demonstrates that the MIMO model learns more precisely the relations between linked domains. UDM appears to be in principle unsuitable for the transfer but for some isolated cases for the translation between lower and upper arm. With regard to the particular transfer method employed, it is difficult to conclude which performs better from the obtained results. It could be said that for the evaluated translations the transfer of activity models in principle provides better results, although the opposite is seen for some other cases.

5.4.3. Transfer between Kinect and IMU

The transfer between sensors of different domain (cross transfer) is here evaluated. The translation between Kinect and IMU relies on the identification of $\Psi_{K\rightarrow I}$ which is next characterized. $\Psi_{K\rightarrow I}$ is a 3-input (3D position) 3-output (3D acceleration) MIMO mapping with 10 tap delays (q = 10, 108 parameters to learn). As for the transfer between IMUs, the tap delay is set to this value to ensure that the MIMO model captures the dynamics of the underlying relation that links the domain transformation.

The specialization of the transfer of activity templates ($\mathcal{T}$) and activity models ($\mathcal{M}$) to the Kinect and IMU test case is now described. The Kinect recognition system is based on $\mathcal{T}$, while the IMUs are seen to employ $\mathcal{M}$. Accordingly, the transfer learning from Kinect to IMUs will make use of the transfer of activity templates whilst the translation from IMUs to Kinect will be performed through the transfer of activity models.

First of all, the signal mapping from Kinect (position) to IMU (acceleration) is identified. The **System Identification** process characterizes through:

- A MIMO mapping $\Psi_{K\rightarrow I}$ from the 3D Kinect joint coordinate, to the 3D acceleration (intuitively it can be seen that this requires

the MIMO mapping to realize at least a 2nd order differentiation) is learned.[8]

- This model can be used both to translate from Kinect to acceleration, and from acceleration to Kinect, thanks to the two transfer models. The reverse MIMO mapping is not needed.

For the **Transfer of Activity Templates (from Kinect to IMU):**

- The source domain recognition system works on the 3D position coordinates. It also stores the activity templates $\mathcal{T}_S = \mathcal{T}_K$ that are the 3D joint coordinates for each gesture.
- $\mathbf{x_S} = \mathbf{x_K}$ is the 3D joint position measured by the Kinect sensor (source)
- $\mathbf{x_T} = \mathbf{x_I}$ is the 3D acceleration measured on the body (target)
- $\hat{\mathbf{x_T}} = \hat{\mathbf{x_I}} = \Psi_{K \to I}(\mathbf{x_K})$ is the acceleration predicted on the body from the known joint position.
- After template translation, $\mathcal{T}_T = \mathcal{T}_I$ are the predicted 3D on-body acceleration patterns for each gesture and the corresponding class labels.
- The target recognition system is automatically trained at runtime on the templates $\mathcal{T}_T$, and finally operates on the acceleration sensed by the IMUs.

For the **Transfer of Activity Models (from IMU to Kinect):**

- The source domain recognition system works on the 3D acceleration sensed by an IMU. It uses models $\mathcal{M}_S = \mathcal{M}_I$ for the recognition.
- $\mathbf{x_S} = \mathbf{x_I}$ is the 3D acceleration measured on the body (source)
- $\mathbf{x_T} = \mathbf{x_K}$ is the 3D joint position measured by the Kinect sensor (target)
- $\hat{\mathbf{x_S}} = \hat{\mathbf{x_I}} = \Psi_{K \to I}(\mathbf{x_K})$ is the acceleration predicted on the body from the joint position.

[8] I or K are used instead of the S or T subscripts in Ψ or $\mathbf{x}$ to be specific about whether the signals come from the IMUs or the Kinect.

- After translation, the 3D joint coordinates of the Kinect are mapped to "look like" acceleration. The recognition models devised for the IMU are used as is by the target system ($\mathcal{M}_T = \mathcal{M}_S$) that now operates on the Kinect data.

According to the devised experimental setup, six cases of translation of AR from a source to target domain are studied. Three are cases of translation of an existing ambient AR system operating on the joint positions delivered by the Kinect hand, towards a system which will use body-worn accelerometers for AR. The on-body sensors are worn either on the lower arm, the upper arm, or the back. This is done with the transfer of activity templates. Three other cases translate an existing wearable AR system operating on on-body accelerometers towards an ambient system which will use the joint position of the hand delivered by Kinect. The on-body sensors are worn either on the lower arm, the upper arm, or the back. This is done using the transfer of activity models. These scenarios are devised to help investigate the potential of the transfer methods for sensor systems of diverse modality (position/acceleration) and of diverse characteristics (centered on close-by or unrelated body parts).

System identification performance

The *BestFit* computed between the acceleration measured at the lower arm, upper arm and back, and the acceleration predicted from the Kinect hand position is presented in Figure 5.7. The best fit tends to be obtained with PDM (Figure 5.7(a)). This may be expected, as the mappings are learned on the dynamics of all gestures. Learning can also occur on one gesture of a given class (Figure 5.7(b)). In that case, the best unique gesture to learn a mapping model appears to be the circle, followed by triangle or square (highest *BestFit*). Thus, one gesture may be sufficient for the mapping to capture the dynamics of the physical system and extrapolate to a wider range of body movements. UDM does not achieve an adequate mapping (Figure 5.7(c)). The idle dataset has only rare occurrences of larger amplitude limb movements and is insufficient to represent the dynamics of the physical system. Nevertheless, a dataset from a domain which does not comprise the activities to recognize but which contains richer limb movements may possibly be used to learn a satisfactory mapping model. In either case, it will be seen later that this can be compensated by using more data (Figure 5.9). The fit worsens for mappings between less related body

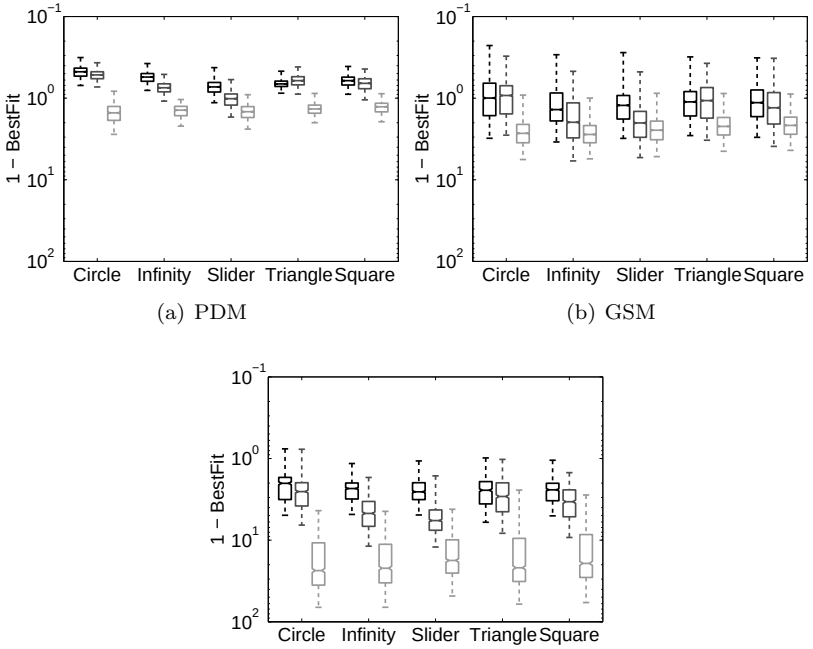
regions. Nevertheless, the fit between hand position and upper arm acceleration is close to that of the hand position to lower arm acceleration. This suggests that this approach may be applicable to transfer activity models across close-by and related limbs also for cross domains. The back acceleration is hardly predictable from the hand position, which is consistent with the explanation given for the IMU to IMU translation to this respect.

Transfer accuracy

Source and target baseline classification accuracies are assessed by training and testing on data from the same domain (position or acceleration). The Kinect to IMU translation is evaluated by training the recognition system on the predicted acceleration $\hat{\mathbf{x}}_{\mathbf{I}} = \Psi_{K \to I}(\mathbf{x}_{\mathbf{K}})$ and testing it on the measured acceleration $\mathbf{x}_{\mathbf{I}}$. The IMU to Kinect translation is evaluated by training the recognition system on the measured acceleration $\mathbf{x}_{\mathbf{I}}$ and testing on the predicted acceleration $\hat{\mathbf{x}}_{\mathbf{I}} = \Psi_{K \to I}(\mathbf{x}_{\mathbf{K}})$.

Figure 5.8 depicts the classification accuracies baselines in the source (BS) and target domain (BT) and those after transfer to the target domain. As well as in the IMU to IMU case transfer here FS2 is utilized. The baselines represent the accuracy obtained by a recognition system that is trained and tested on the same sensor (no transfer). The GSM mapping is learned on the "circle" gesture, which was identified as the best gesture to learn a mapping model. Similarly as for the IMU to IMU scenario, the baselines indicate that the gestures can be classified with an accuracy of 98% or more with the lower-arm acceleration and the upper-arm acceleration. Similar baseline performances are also obtained for the hand position registered through the Kinect sensor. The results after transfer must be assessed with respect to the performance drop from the baselines. In this regard, the performance drop from BS indicates how much worse the system becomes after transfer while the drop from BT indicates how much better would be a system devised specifically for the target domain.

In the transfer between hand position and lower or upper arm acceleration, the PDM and GSM models tend to perform equally well. The best results are obtained when translating from hand position to lower-arm acceleration or vice versa, with less than 4% drop from BS. The drop in performance from BS is less than 8% for the translation from hand position to upper-arm acceleration and vice versa. The direction of the transfer does not affect the results much. The GSM results show



(a) PDM (b) GSM

(c) UDM

Figure 5.7: *Logarithmic box plot. The boxplot represents the statistic distribution of the sample set (the central mark is the median, the edges of the box are the 25th and 75th percentiles, and the whiskers the most extreme results not considered outliers) of* $1 - BestFit$ *between the acceleration measured at the lower arm, upper arm and back (first, second and third box within each gesture group) and the acceleration predicted at that location from the position of the hand measured by Kinect. (**a**) The mapping is trained on all gestures and the fit computed on the indicated gestures. (**b**) The mapping is trained on the indicated gesture and the fit computed on all of them. (**c**) The mapping is trained on data from another domain, and the fit is computed on the indicated gestures.*

that executing a single "circle" is sufficient to identify a mapping model that leads to a transfer with performance drop between 1% to 7% from BS. The transfer between the hand position and the back acceleration shows a large drop from BS with all mappings (30% to 70%). This is

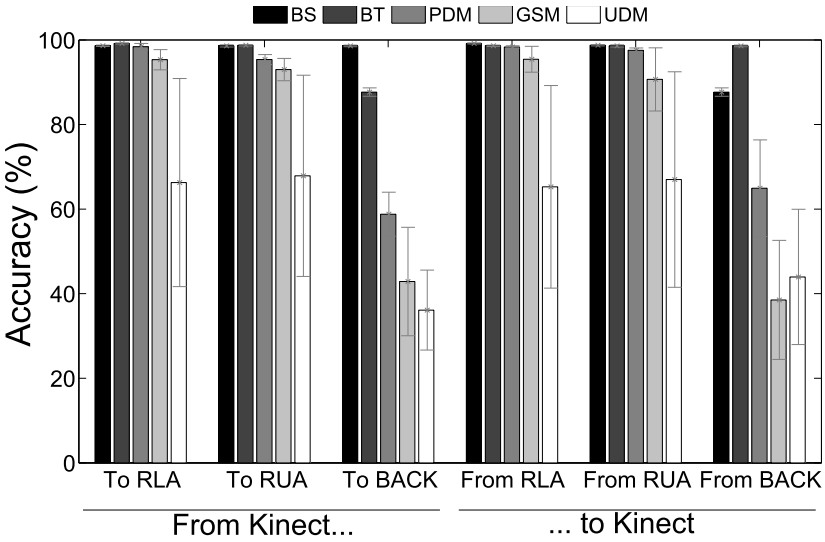


Figure 5.8: *Classification accuracy (average - bar - and standard deviation - whiskers -) for the translation between an ambient and wearable system with FS2. Left half: transfer from a system trained on the Kinect hand position to a system operating on the acceleration measured at the indicated positions. Right half: transfer from a system trained on the acceleration signals measured at the indicated positions to a system using the Kinect hand position. BS and BT indicate the baseline accuracies obtained with a system trained and tested on the source and target domain.*

consistent with the low *BestFit* obtained when attempting to predict the back acceleration from hand position. UDM appears to be in principle unsuitable for the transfer. This is consistent with the analysis of *BestFit*.

The UDM mapping improves when learned on more "idle" data (Figure 5.9). With 2000 samples (67 seconds), the performance is about 15% to 30% below the corresponding baselines for FS2 and FS1 respectively. This indicates that, with sufficient data, a dataset from an unrelated domain allows the MIMO mappings to capture the dynamics of the physical system. The difference between FS1 and FS2 highlights that an automatic selection of better features by the source or target system may lead to improved results. Thus, the reported results are a

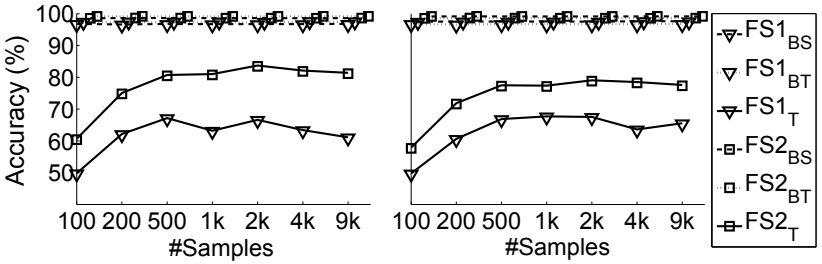


Figure 5.9: *Effect of the amount of idle data used to learn the UDM mapping on the translation accuracy from the Kinect hand position to acceleration at the lower arm (left) and vice versa (right) for feature sets 1 an 2 ($FS1_T$, $FS2_T$). BS and BT are the source and target baselines.*

lower bound on the performance.

5.4.4. Discussion

Instruction of newcomer sensors

Real-world AR systems are subject to changes in the sensor setup normally due to equipment upgrades and maintenance (i.e., replacement or addition). To make use of the newly introduced sensors, standard AR systems normally need from a complete retraining. This training requires the collection of large datasets which turns to be completely unsuitable in real-world applications. The transfer approach proposed in this work serves to translate the recognition capabilities of an existing system to a newly introduced untrained system, which may be performed at runtime and without requiring expert or user intervention.

Although this study has particularly explored the transfer between recognition systems operating on individual sensor nodes, approaches such as the ones suggested in previous work of this thesis may easily benefit from this. Concretely, the HWC provides a scheme on which each sensor node operates in an independent basis, thus it is completely feasible to only train the corresponding sensor classifier associated to the newcomer node, without requiring to stop the recognition process at any rate. Thus for example, the HWC may help provide activity-awareness until the replacement for a broken sensor is provided. Assuming that the new sensor is of different characteristics to the substituted one, a new sensor classifier must be developed for the newcomer

system. Similarly, the HWC supports the addition of new nodes at the network level, which also require to build new sensor classifiers. The training of this sensor classifiers may be performed in a simple and agile manner through the transfer methods proposed here.

Transfer method advantages

The evaluation of the proposed approach has been demonstrated to succeed for transfers between same and different domains. The model is capable of capturing the underlying relation between systems of identical modality translated or rotated with respect to each other. Similarly, the proposed transfer approach proves to be capable of discovering the physical relation between cross domains. Although this has been here demonstrated for Kinect-based and accelerometer-based systems, the approach itself is generic and can be applied to other sensing systems.

The method should scale well with the number of classes, since a mapping learned with one instance of a single class (GSM) performed well on the prediction of the signals of other gestures. This indicates that the mapping approximated the physical relations between the sensing systems, independently of the gestures. Isolated AR was in this work evaluated, but the approach is also applicable for continuous recognition (spotting). Therefore, once a mapping model is identified, the target signal can be transformed to essentially be like a signal from a source system on which an existing spotter operates.

Low-variance data unrelated to the activities of interest can be used to learn a mapping (UDM), albeit with more data. This has practical benefits, since "unrelated" domain data can easily be acquired "in the background", whenever the user is in the sensing range of source and target sensor systems. For example, when a new sensor is added or replaced, the learning may be performed while the user executes their daily routines, yet in a reasonable time period. Learning, however, benefits from the execution of movements highlighting the physical relation between the sensor systems.

The learning of the mapping model is not affected by the choice of a feature set. This work analyzes the system performance with predefined simple feature sets, which suffice for the AR problem complexity. In either case, it is proved that this constitutes a lower performance bound of the system. In the model transfer architecture the target system may run a feature extraction and selection method, and thus possibly improve upon the results presented here. Assuming an ideal mapping

model, the signal transfer architecture would also benefit from feature extraction and selection in the source domain.

The approach was evaluated in simulation. It is however entirely suitable for online, real-time implementation, for instance in sensor nodes.

Transfer learning principles allowing a trained system to transfer activity recognition capabilities to another system have been proposed for body-worn sensors [93] and ambient sensors [85]. These approaches have in common that they operate on long time scales as they require all the relevant activities to be observed several times (e.g., timescale of days or more). Moreover, the transfer may be incomplete when part of the activities are not performed. These models present some other important limitations such as the need of predefining allowed run-time variations or not be defined for adaptation across sensor modalities. The approach proposed here neatly overtakes previous contributions by fulfilling these requirements of real-world AR systems. The proposed transfer model can be performed in a very short time scale, it is capable of learning despite sensor variations such as rotations or translations, and also applies well to cross domain transformations.

This approach may be useful in crowd-sourcing scenarios [137] to translate generic activity models to the specific sensor modalities that one user has. Moreover, some fields such as gaming could highly leverage from this type of learning since multi-sensing platforms not envisioned during design time may be easily learned and utilized to enhance user experience. For example, users may play with the Kinect when staring in front of the camera, but keep playing when they get out of the sensing range of this sensor through transferring recognition capabilities to a smartclock or smartphone. Transfer learning may also facilitate games developers tasks when porting functionality and playability from a sensing platform (e.g., Kinect) to a different but related sensing equipment (e.g., Wii Remote or Motion Plus).

System identification

For specific and tractable cases, a white-box mapping may be devised. Nevertheless, that approach would not generalize to modalities or configurations not foreseen by the system designer (changeable configurations). The approach that is here proposed allows us to take advantage of additional sensors as they become available and to learn the mapping without expert intervention. The approach may be improved by

non-linear or time-varying models, e.g. with time-delay neural networks [120] or non-linear ARMA [113], or by modeling the transformations between multiple sensors (e.g. two joint coordinates and one acceleration). These models support non-linear transformations, such as those needed when a sensor changes properties over time. Nevertheless, more complex transformations likely need longer coexistence time between source and target to estimate the model parameters. Besides, non-linear models are more prone to overfitting and learning of noise and signal artifacts which typically appear on the registered data. The possibility of updating the system identification model instead of relearning it from scratch is also of much interest in our context. On-line non-linear system identification models are meant to do that [117].

Gray-box (GB) models, an intermediate approach between BB and WB modeling could be also considered. GB models combine the best of WB and BB approaches [138], this is, knowledge-based modeling through mathematical equations that describe the physical process and black-box modeling, whereby a parametrized model is designed and whose parameters are estimated solely from measurements made on the process. Yet, the limitations already described for WB may similarly apply to GB, thus the extent to which these models may be applicable in our context depends on how the GB model is specifically defined.

The mapping between the modalities considered here might look trivial at first glance. For example, accelerations of the body are related by biomechanics constraints, and acceleration can be calculated from position. However, the acceleration measured by the IMUs is in a local, time-varying frame of reference. The frames of reference of the sensors are not identical and their relative orientation may change over time. Moreover, these modalities exhibit complex transfer characteristics between each other. In fact, there is a dynamic relation between acceleration and position involving current and past samples. Thus, this setup allows us to conduct a non-trivial, yet manageable analysis of our method.

System identification models are shown to cope with the complexity of the signal mapping. Nevertheless, it must be present the importance of synchronization among source and target signals during the mapping learning process. Although the models have a certain tolerance (not analyzed here) a significant delay or jitter might prevent us from obtaining a functional transfer model. Therefore, synchronization among signals as well as (re/down)sampling techniques should be somehow considered at the point of need.

BestFit

The *BestFit* indicates the quality of the mapping and to some extent the quality of the resulting translation. This may be an indicator guiding the self-organization of an ecology of sensor systems for opportunistic AR [139].

Since the *BestFit* tends to be highest when sensors measure the movement of the same limb, further investigation may evaluate whether this could be used to automatically localize on-body IMU placement when in range of a skeleton tracking system. The use of other statistical metrics such as mutual information are also here identified for future work in this direction.

Transfer of Activity Templates vs. Transfer of Activity Models

The transfer models differ in their computational complexity and memory needs. Templates transfer does not add computational load on the target system after translation but it requires the source system to store activity templates. However, this does not demand large amount of space (47kBytes in floats here). This is well suited for an ambient source and a wearable target system. In contrast, transfer of activity models requires that the target sensor signals are continuously translated. This increases the computational load on the target, but the mapping complexity is low and easily benefits from single instruction multiple data (SIMD) computation. The storage requirement is lower, since only the activity models need to be stored. This is well suited for a wearable source and an ambient target system. The models also differ in whether the source signal is mapped to the target, or vice versa. If a mapping model exist both ways, as it happens to occur for the IMU to IMU case, then the choice of the transfer is based on computational and memory requirements. If the mapping model is more accurate in one way, then the transfer that uses this mapping is favored. In this work the mapping from position to acceleration influenced the transfer model choice.

Supporting adaptive and evolvable AR

Sensor setup changes do not only appear after a maintenance or upgrade operation. In fact, different sensor configurations are envisioned during the course of a user's normal day. Depending on the particular context, users may wear specific garments (e.g., at work), casual

clothes (e.g., at home) or specific accessories (e.g., at the gym). During these situations, sensors may be removed, substituted or newly added. Classic AR systems are trained on sensor data streams from datasets collected at design time with predefined and optimal sensor configurations. Then, accounting for these variations would require here again to collect as many datasets as possible sensor configurations, which happens to be unfeasible. Transfer approaches are devised as a means to help support a continuity of recognition by adapting to the actual sensing configuration. AR systems may dynamically transfer their recognition capabilities among themselves, thus all available systems may be readily prepared to be used for the required activity-aware services.

The potential of transfer learning is not only restricted to wearable sensing but other sensing platforms. In fact, an important feature of activity recognition systems is to provide a continuity of context-awareness across different sensing environments, as the user changes location or carry-on devices. As the user performs their daily activities, various sensor systems may be discovered. These sensor systems may not necessarily be capable of activity recognition, as they may also be deployed for other purposes. For instance, a user relies on a smartphone for activity awareness (e.g. for energy expenditure analysis). The user enters a room with an activity-aware gaming system and leaves the smartphone on a desk. The smartphone now cannot recognize the user's activities. The gaming system can sense their movements, but may not be devised to recognize the same activities as the smartphone did. Thus, in principle, even if the gaming system sensors deliver relevant data, these data cannot be used to substitute the phone sensors, unless some translation occurs. The use of the proposed transfer methods may allow to provide the gaming system with the smartphone recognition capabilities. This kind of situations, which fit well with the experimental setup evaluated in this work, are seen to be more and more frequent due to the increasing sensor equipment of living environments and users.

Transfer learning is observed to be very important to keep the knowledge gained during the use of the recognition systems. AR systems should be defined to personalize and evolve while learning from the user experience, context and conditions. All this knowledge may be potentially lost when the system is forced to operate on new sensing infrastructure. The use of transfer learning techniques such as the proposed in this work may help to translate this knowledge to the new domain, thus allowing the system to function in a natural way.

Open issues

In the current dataset, the gestures are performed in a quite slow way, leading to accelerations with intensity lower than $1.3g$ for most cases. This means that the accelerometer readings are mainly due to the time-varying angle formed between the sensor and the gravity vector. If higher dynamics are present, the mapping has to learn how to calculate the transformed static part (due to gravity) and the dynamic part (due to movements) and the transformations are not necessarily identical (e.g., accelerometers mounted on two different positions of the forearm might have the same orientation with respect to gravity, but the dynamic acceleration will in principle grow with the distance from the elbow). Furthermore, since accelerometers alone measure their data within a local frame of reference, while an external positioning system (like Kinect) collects data referred to a fixed (world) frame, the rotation matrix needed for transformation is dependent on the body posture itself in a non-linear manner. In other words, the signal mapping would have to include not only a second derivative, but also a rotation which is depending non-linearly on the body posture. The linear MIMO model can only approximate the second derivative and a fixed rotation, which would be an average rotation. This may become an issue with more ample movements, but in our dataset the relative rotation of the frames of reference was limited for most gestures ($\pm$30-40°). Only for the slider gesture the lower arm rotates by almost 90° at the extreme of the movement, compared to the starting position.

Kinect and other video-based tracking systems are affected by occlusions. Since only a small amount of data are needed to learn the mapping, this process is likely feasible in-between occlusions. Furthermore, during an occlusion the *BestFit* decreases, so it may be enough to let the system learn only when *BestFit* is higher than a certain threshold.

Another limitation is that some sort of movements may not be sensed by certain modalities. For example, Kinect cannot detect torsions of hand and forearm (e.g., in gestures like turning a knob or tightening a screw). These torsions translate into changes of the local IMU frame of reference with respect to the camera world coordinates by Kinect. Figure 5.10 (left) shows an ilustrative example where the arm is rotated along its generation axis but no movement is appreciated from the body joint camera model. This may be part of an specific gesture or movement as shown in Figure 5.10 (right). In both cases diverse ac-

celeration signals may normally be expected (at least variations on the static component resulting from the IMU frame of reference rotation). In contrast, torsions are easily sensed by gyroscopes and accelerometers, meaning that the expected transfer performance is modality- and gesture-dependent. This makes this approach well suited to opportunistically improve the mapping models by taking advantage of additional sensor modalities as they become available.

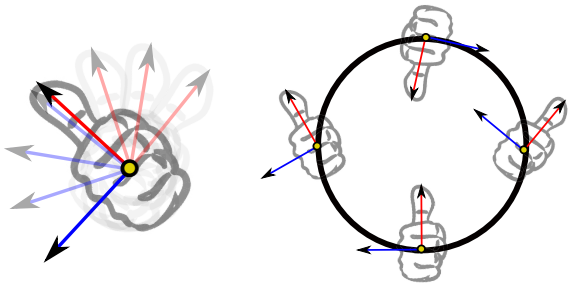


Figure 5.10: *Example of torsion along the forearm (left). Example of torsion when performing a gesture (right). The superimposed axes show the change of the local frame of reference of an IMU sensor placed on the wrist. Torsion may not be sensed through the 3D-position Kinect model.*

The problem of having diverse possible representations for a gesture or movement in one domain that may correspond to a unique representation in the other domain also applies the other way around. Let us imagine a subject that performs a gesture in a given position, and then moves to another distant position (yet in the range of Kinect) and executes an exact reply of the very first gesture. The registered inertial signals are seen to be identical for both executions but different for Kinect. A small set of tests was performed to analyze how the mapping model deal with this situation. First, all the Kinect 3D coordinates of the hand signals where referred to a similar origin randomly selected. No significant difference was encountered in terms of both fitness and transfer accuracy. Moreover, it was evaluated the removal of the body model center of mass to provide a common reference. Again, no better results were obtained than for the original data. From a mathematical point of view this is expected since the derivatives (position to acceleration) filter out the offset between both spacial gesture realizations. This confirms once more the capacity of the considered mapping model

for capturing the physics of the underlying systems relation.

Kinect suffers from other more complex technical limitations on which the mapping model is likely to fail. One example is the abnormal skeleton model obtained when the subject interacts with some particular objects. When the user holds an object like a stick or gets too much close to other habitual items of the environment such as tables or chairs, Kinect tends to incorporate these objects as part of the complete skeleton model (e.g., a stick may be interpreted as an extension of the hand). These and other previously presented problems are currently being investigated, for example through the use of arrays of cameras or techniques based on the body velocity measurement to avoid the occlusions [140, 141].

5.5. Conclusions

AR systems of the real-world are subject to sensor setup changes. Obsolete or damaged sensors may be replaced with sensors of different characteristics or new sensors added as part of equipment maintenance and upgrades. Moreover, sensors availability may also vary during the normal course of a person's day depending on their particular context (e.g., a user wears specific apparel at work, casual clothes at home or fitness accessories at the gym). These sensors may not be capable of specific activity recognition since either they may not have associated activity models or are originally devised for other purposes. However, most of this sensing equipment could be profit for AR purposes since they are in principle capable of measuring human behavior (e.g., a bracelet to detect dietary conducts could be also employed to monitor the user workout). In all cases, a specific training is required for these newcomer systems to become usable for the recognition of the activities of interest. To build the recognition models, the collection of new experimental data is in principle required, which happens to be unpractical and unapproachable in realistic scenarios. Transfer learning is here devised as the perfect medium to perform this training in a suitable way.

In this work, a means to automatically translate AR systems between sensor modalities, thereby effectively allowing to transfer AR capabilities from an existing or source system to an untrained or target system has been proposed. The approach relies on the learning of a mapping between source and target sensor signals. Two transfer modes

are proposed for the translation of AR systems that operate either on activity templates or activity models.

The transfer learning models are evaluated in a multimodal gesture recognition setting consisting of a vision-based skeleton tracking system and body-worn accelerometers. For this, two scenarios of transfer learning have been analyzed. For the first case, the transfer between sensors of the same modality (identical transfer) is studied. To that end, the translation between accelerometers placed in different body parts is evaluated. System identification techniques can be used to learn a linear MIMO model that maps the 3D accelerations sensed by source and target IMUs. As few as a single gesture (3 seconds) of data is enough to learn a mapping model that captures the dynamics of the physical system. This permits to perfectly transfer the recognition capabilities of systems that operate on close-by or related sensors. The IMU to IMU translation across adjacent limbs (here lower arm and upper arm) achieves a recognition accuracy superior to 97%, which is less than 2% below the accuracy of the initial system. This proves to be independent of the direction of the transfer.

The second scenario refers to the case in which the transfer between sensors of different modality (cross transfer) is pursued. A linear MIMO model that maps 3D positions sensed by a Kinect to the 3D acceleration measured on-body by IMUs can be obtained through system identification. Here again, a sole gesture is sufficient to learn a mapping model that captures the dynamics of the physical system and generalizes to unseen movements when the user is active. When the user is idle, more data is required to learn this mapping, yet much less than other transfer approaches. The Kinect to IMU and IMU to Kinect translation achieves a recognition accuracy of 95%, and is less than 4% below the accuracy of the initial system. When translating across sensor modalities and also to an adjacent limb (e.g., Kinect hand to IMU on the upper-arm), the accuracy after translation is 8% below baseline.

The approach is generic and could be applied to other sensors, for example between a gyroscope and an angle sensor such as a stretch sensor integrated in clothing. The approach is suited for online use and can be similarly used to translate gesture spotting capabilities across modalities.

The MIMO models can be replaced by nonlinear ARMA models [113], time-delay neural networks [120] or evolutionary techniques [142], that may help capture more complex dynamics of the physical system, for instance for combinations of sensors.

This work supports the translation of AR capabilities between sensor modalities without user or system designer's intervention. This is an important characteristic for AR in open-ended environments, where the nature and availability of sensors may change over time.

The approach supports the multi-modal recognition of activities by allowing for example to combine video and motion information. Computer vision and wearable inertial sensing are probably the most prolific domains in AR. Connecting these two domains is encountered to be of high value to the field. This work contributes to define a means to relate both domains and even transfer AR models that are predefined for operating in one domain to be used in the other domain. In the future, this may be further used to learn activity models from existing annotated video sources (e.g., from YouTube), and apply them to movement data sensed on the body (e.g., with a smartphone). This is supported by recent results in skeleton tracking from monocular cameras [143].

6

Conclusions

6.1. Achievements

Much research has been performed during the last years in the activity recognition (AR) domain. Nevertheless, most of the activity recognition solutions provided so far are devised for ideal scenarios, thus they may unexpectedly behave when applied to real-world conditions. Therefore, the goal of this thesis was to investigate on the effects of some of the most prominent technological and practical challenges posed by the use of on-body inertial sensing human activity recognition systems in the real-world. In the following, the achievement of the three objectives defined to support this goal is described.

Objective 1: Investigate the tolerance of standard AR systems to unforeseen sensor failures and faults, as well as contribute with an alternate approach to cope with these technological anomalies.

Classic AR systems assume that the sensor setup remains invariant during the lifelong use of the system. Nevertheless, as other electronic devices, on-body sensors are subject to faults and failures. These technological anomalies lead to changes in the sensor data streams, which are normally unforeseen during the design phase and unpredictable at runtime. Consequently, models trained on ideal signal patterns may react in an undesired manner to anomalous sensor data. This potentially translates into a partial or total malfunctioning of the AR systems.

In this work, the principal sensor technological anomalies have been introduced and particularly categorized in failures and faults. For the first of these categories, a qualitative evaluation of the effects of sensor failures on standard AR systems has been performed. From here, systems based on a single sensor have been demonstrated not to operate when the sensor stops delivering data. The use of multi-sensor setups has been proposed in the literature to overcome this limitation through leveraging the remaining active sensors. However, not all sensor fusion models have been proved to be capable of dealing with failures. In fact, multi-sensor fusion at the feature level has been shown not to function when a sensor fails. Multi-sensor decision fusion models have been used in related work to deal with sensor breakdowns since independent recognition systems are defined for each sensor node. Nonetheless, hierarchical decision and majority voting, two of the most widely used decision fusion methods have been here demonstrated to present important limitations to overcome the effects of sensor failures. Taking

into account the shortcomings of hierarchical decision and majority voting models, a novel methodology that combines their principal advantages and avoids their main limitations has been proposed in this work. This innovative model, so-called hierarchical weighted classifier (HWC), has been benchmarked in ideal circumstances and proved to provide a performance comparable to the standard AR systems. More importantly, the HWC has been proven to succeed when dealing with the effects of sensor failures, even when a majority of the sensors fail. For the second category of sensor technological anomalies, namely sensor faults, the tolerance to the effects of sensor dynamic range changes has been evaluated for the proposed HWC model and standard AR approaches. Systems of a single sensor and those of multi-sensor fusion at the feature level have been demonstrated to severely degrade their recognition performance under harsh conditions. State-of-the-art decision fusion models have proved a higher fault-tolerance, but still far from being capable of providing accurate activity-awareness. Conversely, the HWC model has been shown to be tolerant to sensor faults when a minority of the sensors are damaged, here serving to provide accurate activity recognition in faulty setups.

The first objective of this thesis has been fulfilled since the behavior of standard AR systems have been evaluated under the effect of sensor technological anomalies and a novel approach has been alternatively proposed to cope with the effects of sensor failures and faults.

Objective 2: Research the robustness of standard AR systems to unforeseen variations in the sensor deployment, as well as contribute with an alternate approach to cope with these practical anomalies.

Most AR systems assume a predefined sensor deployment that further remains unchanged during runtime. Nevertheless, these are not lifelike assumptions. During the normal use of the systems, users may place the sensors in a different position to their ideal or default distribution. Likewise, the sensors may move from their original location to a different one due to a loose-fitting attachment. AR systems learned on activity patterns characteristic of a default or ideal deployment may likely incur in misrecognition due to sensor displacements.

In this work, the concept of sensor displacement has been presented and particularly defined as a combination of rotations and translations. Moreover, the principal effects of sensor displacement on on-body inertial modalities have been illustrated. To investigate the practical con-

sequences of sensor displacement two approaches have been followed. Firstly, the effect of sensor displacement on AR models has been analyzed for synthetically modeled rotations and translations. From here, recognition systems based on a single sensor have proved to be very sensitive to both rotations and translations, specially when the displaced sensors are originally devised to be mounted on body parts subject to low mobility. Multi-sensor configurations have shown more tolerance to sensor deployment variations. Feature fusion approaches have only demonstrated moderate robustness for subtle deployment variations, as well as being incapable of recognition when more than one sensor is displaced. The innovative HWC model proposed in this thesis has been proved to satisfactorily cope with slight to moderate variations for any number of displaced sensors, and to be also good enough when a minority of the sensors are hardly modified with respect to their original positions. Secondly, the effects of sensor displacement have been investigated when observed in realistic settings. As part of this work, a novel benchmark dataset has been collected since there was none available for the study of this kind of problem. This dataset has been devised to innovatively explore sensor displacements generated by people when self-attaching the devices and large sensor depositions purposely introduced to represent boundary conditions for recognition algorithms. For this dataset, a statistical analysis of the effects of displacement has been performed to quantitatively demonstrate the variations introduced with respect to the ideal sensor deployment. A relevant disparity in the data has been observed when the user self-placed the sensors, which substantially increased when a large sensor depositioning was intentionally introduced. The effect of these variations on the AR systems performance has been further analyzed. Neither single sensor recognition systems nor multi-sensor feature fusion approaches have demonstrated to be capable of coping with sensor displacements motivated by the user self-placement or the large depositioning of the sensors. The HWC model has conversely proved outstanding capabilities to assimilate the changes introduced during the self-placement of the sensors and to moderately overcome the situation of largely depositioned sensors.

The second objective of this thesis has been fulfilled since the behavior of standard AR systems have been evaluated under the effect of sensor deployment variations and a novel approach has been alternatively devised to deal with the changes introduced by sensor displacements.

Objective 3: Study the capacity of standard AR systems to support unforeseen changes in the sensor network, as well as contribute with an alternate approach to cope with these topological variations.

Real-world AR systems are subject to sensor setup changes. During equipment maintenance obsolete or damaged sensors may be replaced with sensors of different characteristics. Likewise, new sensors may be added as part of system upgrades. Moreover, sensors availability may vary during the normal course of a person's day depending on their particular context. In these scenarios, newcomer sensors may not be capable of recognizing specific activities because they do not have associated the required activity models. Consequently, a specific training of new AR models is required to become usable for the recognition of the activities of interest. To build the recognition models, the collection of new experimental data is required, which happens to be unpractical and unapproachable in realistic scenarios.

In this work, it has been analyzed the shortcomings of transfer learning approaches devised in the literature as practical solutions for training newcomer sensor systems. Taking into account these limitations, it has been proposed a novel means to automatically translate AR capabilities from an existing or source system to an untrained or target system even for different sensor modalities. This new approach, which relies on the learning of a mapping between source and target sensor signals, has been defined for the translation of AR systems operating either on activity templates or activity models. The proposed transfer method has been evaluated in a multimodal gesture recognition scenario consisting of wearable and ambient sensors. Transfers of AR systems between sensors of the same modality (identical transfer) and between sensors of different modality (cross transfer) have been assessed. The identical transfer has been here evaluated for the translation between IMU-based systems. In this case, the translation of activity templates and activity models across sensors placed on adjacent limbs has been proved to perfectly succeed, thereby maintaining nearly the same performance as the source system. Likewise, cross transfers, here defined for the translation between IMU and video-based systems, have been shown to provide a quite similar performance for the translations between contiguous limbs. For both identical and cross transfers, the data of a sole gesture has been proved to be sufficient to learn a mapping model that captures the dynamics of the physical underlying relation between source and target systems. This mapping learned for a single

gesture has been demonstrated to generalize to other unseen movements. More data has been determined to be required to learn this mapping when the user behaves arbitrarily or idly. Yet, much less data is required than for other state-of-the-art transfer approaches. Both for identical and cross transfer cases, the translation has been shown to be accomplished in a very short time and without requiring from user or system designer's intervention. These characteristics of the presented method fit in well with real-world AR requirements.

The third objective of this thesis has been fulfilled since the limitations of classic AR approaches to train newcomer sensor systems have been identified and a novel approach has been alternatively proposed to practically learn new AR systems under realistic conditions.

6.2. Contributions

In Section 6.1 it has been proved that the objectives of this thesis have been thoroughly fulfilled. Now, the main contributions of this thesis are listed:

- Identification of the requirements and challenges posed by AR systems in real-world conditions.
- Evaluation of the tolerance of standard AR systems to sensor technological anomalies, particularly sensor failures and faults.
- Definition and development of a novel model, so-called HWC, to overcome the effects of sensor failures and faults.
- Evaluation of the robustness of the proposed HWC model to the effects of sensor failures and faults.
- Evaluation of the tolerance of standard AR systems to sensor deployment variations, particularly static and dynamic sensor displacements.
- Evaluation of the robustness of the proposed HWC model to the effects of sensor displacements.
- Definition, development and validation of a novel multimodal transfer learning method that operates at runtime, with low overhead and without user or system designer intervention.

- Collection and curation of a benchmark dataset to investigate the effects of sensor displacement, introducing the concept of ideal-placement, self-placement and induced-displacement. This dataset includes a wide range of physical activities, sensor modalities and participants. Apart from investigating sensor displacement, the dataset lend itself for benchmarking activity recognition techniques in ideal conditions. The dataset is publicly available to the research community at http://www.ugr.es/~oresti/datasets.

- Collection and curation of a multimodal dataset to investigate transfer learning among ambient sensing and wearable sensing systems. The dataset could be also used for gesture spotting and continuous activity recognition. The dataset is publicly available to the research community at http://www.ugr.es/~oresti/datasets.

6.3. Outlook

Given the novelty of this work, there is still much room to investigate new methods and approaches. In this section, possible future directions to continue and extend the work presented in this thesis are described.

6.3.1. Collection of large standard datasets

One of the main limitations of the wearable AR domain refers to the lack of available datasets to benchmark new models and compare them with prior work. This demonstrates of key importance in a field that lacks of a specific gold standard. Consequently, a strong effort must be put by the wearable scientific community to collect new datasets that may serve to validate AR approaches. In the domain of real-world activity recognition, new realistic datasets should be collected to observe some of its associated challenges. Thus for example, datasets addressing real-life sensor faults and failures are seen to be of much value to evaluate the capabilities of models dealing with technological anomalies. Likewise, an extension of the dataset proposed here to investigate the effects of sensor displacement would be appreciated. Concretely, it would be interesting to observe the displacement introduced during the normal use of instrumented gadgets or garments deployed in realistic scenarios. This would reduce the influence on people's normal behavior because of their experimental awareness. The use of massive- and

crowd-monitoring protocols is also seen as a leading direction to enrich datasets, thus allowing for a better understanding of human behavior and people interaction.

6.3.2. Dynamic reconfiguration of the HWC

One of the most relevant properties of the proposed HWC model is its flexibility to sensor setup changes. Sensors may be removed or newly added to the original sensing ecosystem while keeping the consistency of the whole recognition process. This property is specially eligible to intelligently leverage those sensors that happen to be available to the user. Depending on the particular application, the use of a specific part of the sensing infrastructure could be preferred. Moreover, this is seen to be of high importance to support the seamless integration of future sensing technologies with already existing ones. This opens-up a new range of opportunities for AR systems moving from constructive to evolvable paradigms.

AR systems are normally devised for a set of particular activities, however, this may change in the course of time depending on the particular user and application needs. Additional activities to the ones devised at design-time may be required when for example a new exercise routine is considered or a workout plan is modified. These changes are not only seen to add new activities but to also remove some of these at the point of need. This is found of special interest to reduce systems complexity and increase their recognition performance, as well as to procure systems personalization to subjects. Schemes based on standard AR models require a complete system rebuilding when the activity set is varied. Likewise, multi-decision or fusion techniques such as adaboost, decision stumps, random forests or other popular ensembles and meta-learners require to retrain the model. Conversely, the HWC supports this kind of reconfigurations since its flexibility does not only apply at sensor level but also at activity level. New base classifiers must only be trained for the newly added activities, and their associated weights computed. If an activity is rather removed, an update of the model weights is just required. Future work aims at evaluating the capabilities of the HWC model to this respect specially for online realistic scenarios.

6.3.3. Self-adaptive HWC

The proposed HWC model has shown an important tolerance to sensor anomalies. Nevertheless, this approach takes into account the potential decisions provided by all considered sensors, thus reducing its robustness when a majority of the sensors are severely disturbed. The definition of an adaptive mechanism to automatically update the sensor classifier weights could be of much value to reduce the impact of those decisions provided by the anomalous sensors. However, for the adaptation process extra knowledge or feedback is required to determine whether or not a system operates properly. Involving users in this process, for example, asking them to notify whether the recognition system is making a detection error or not, may be burdensome and lead to misreports. Instead, the use of distance measures and information theory techniques to identify erroneous measurements in a multi-sensor setup could be employed. Likewise, the knowledge of the collectivity may be utilized to define an oracle that helps to identify misperforming sensors. These techniques may allow for an autonomous self-adaptation of the HWC model parameters to the actual network conditions.

Certainly, much benefit could be obtained through reducing the weights of the anomalous sensors or even removing them from the decision fusion process. However, sensor anomalies must not necessarily degrade the recognition of all activities. In fact, some activities may be more affected by sensor faults than others as it was shown in this thesis. Likewise, after sensor displacement the signal characteristics of some activities may change to an extent they are not recognizable anymore whilst others are still recognizable. Bearing in mind the flexibility of the HWC structure, the weights of the misrepresented activities could be reduced while maintaining the original weights for the rest of unaffected activities. This is more precise than reducing the weights at the sensor level and permits to leverage the potential of the affected sensor for the well-recognized activities. The self-adaptation of the HWC model through using the aforementioned ideas is aimed to be part of future work.

6.3.4. Tolerance to other sensor technological and topological anomalies

Apart from the sensor anomalies investigated in this work there exist others that would be interesting to be addressed. For example, some sensor technological anomalies are more dependent on how the signals

are simultaneously registered and processed. In multi-sensor configurations diverse signals are recorded through the different sensor nodes. Although assuming the data are collected at a similar sampling rate, drifts among the signals could appear. To avoid these time deviations, wired configurations are normally required which are impractical in realistic scenarios. In wireless approaches synchronization is pursued by using a common clock reference, however, time skews are normally present. The impact of sensor signal drifts also tends to increase as the number of sensors grows. The evaluation of signal drifts in multi-sensor systems (specially for highly dense deployments) is thus seen as an interesting challenge to be addressed in future work.

The effects of sensor displacement have been particularly analyzed for the acceleration domain. Investigating the effects of realistic sensor displacement on other sensor modalities or domains is also found very interesting. In fact, in a recent preliminary work [144] the effects of sensor displacement were researched for various inertial sensing modalities. Future work in other sensor modalities could also include the evaluation of the effects of extreme sensor depositionings. Extreme sensor depositioning refers to the displacement introduced when two or more sensors are exchanged. Thus, they are relocated in body parts completely unrelated to the ones devised at design-time. Although this is less frequent, a user could misplace a sensor during the self-placement process as a consequence of a mistake (e.g., the wrist-sensor is positioned on the ankle). Real-world AR system should also account for these extreme sensor misplacements, thus contributing to a freer use of the devices.

6.3.5. Multiple trainers and complex modalities in transfer learning

In this thesis the mapping between wearable sensors and between wearable and ambient sensors has been analyzed. Concretely, acceleration to acceleration and position to acceleration mappings have been characterized. Although this is a difficult problem *per se*, more complex sensing mappings may be envisioned. For example, ambient sensors like switch buttons could be used to detect household activities. The mapping from this ambient sensing modality to an on-body inertial modality could be normally discarded from a rational perspective. However, through more complex mappings it could be possible to effectively relate them. In this line, the combination of diverse source sensing modalities to predict a given target modality is meant to provide a more complete descrip-

tion of the phenomena. This may potentially lead to more accurate mappings but at expense of a more complex modeling.

Modeling the transformations between multiple sensors (e.g., from two joint positions to one acceleration) is another interesting future research direction. This thesis already demonstrated the need of monitoring diverse body parts to more accurately recognize human behavior. Likewise, the use of multiple sensors is here envisioned to improve the quality and robustness of the mapping among sensor modalities. In this work, the learning of the mapping was performed between individual body joints; however, it is reasonable to think that more accurate mappings may be obtained when using multiple sensors from which the orientation of the complete limb could be deduced.

Benefiting from more sensor inputs and more complex sensor modalities would translate into the need of more sophisticated models. The dynamic selection of the best set of trainers in terms of both sensor characteristics and modalities could be used to provide an optimal mapping model. The definition of optimal sophisticated models opens-up a new spectrum of research opportunities for lifelong AR systems.

7

Conclusiones

7.1. Logros

Son muchas las investigaciones desarrolladas en los últimos años en el área del reconocimiento de la actividad humana. Sin embargo, la mayoría de las soluciones propuestas hasta la fecha en este ámbito han sido diseñadas para condiciones ideales, por lo que pueden comportarse de forma inesperada cuando se aplican a escenarios del mundo real. Es por esto que el propósito fundamental de esta tesis ha consistido en investigar los efectos de algunos de los retos técnicos y prácticos más destacados derivados de la utilización de sistemas de monitorización inercial del movimiento corporal en el mundo real. A continuación se describen los logros obtenidos para cada uno de los objetivos planteados como parte de esta tesis.

Objetivo 1: Investigar la tolerancia de los sistemas estándar de reconocimiento de la actividad ante posibles fallos y defectos imprevistos de los sensores, así como contribuir con una solución alternativa para lidiar con estas anomalías tecnológicas.

Los sistemas tradicionales de reconocimiento de la actividad presuponen que la configuración de los sensores permanece invariante durante toda la vida útil del sistema. Sin embargo, como otros dispositivos electrónicos, los sensores vestibles experimentan fallos y defectos. Estas anomalías tecnológicas introducen variaciones en las señales normalmente registradas a través de los sensores, siendo imprevisibles durante la fase de diseño e impredecibles en tiempo de ejecución. Por consiguiente, los modelos entrenados con patrones de señales ideales pueden reaccionar de una forma indeseada al recibir datos anómalos por parte de los sensores. Esto se traduce en un potencial mal funcionamiento parcial o total del sistema de reconocimiento de la actividad.

En esta tesis se han presentado las principales anomalías tecnológicas que pueden sufrir los sensores, y se han categorizado en errores críticos o fallos (ej., rotura por caída) y anomalías parciales o defectos (ej., bajo nivel de batería). Para la primera de estas categorías se ha realizado una evaluación cualitativa de los efectos de los fallos de los sensores en los sistemas estándar de reconocimiento de la actividad. Se ha discutido la obvia inoperancia de los sistemas basados en un solo sensor cuando este deja de transmitir datos. Para superar esta limitación en la literatura se ha propuesto el uso de configuraciones multi-sensor que aprovechan los restantes sensores que permanecen activos a través

del uso de mecanismos de fusión. Sin embargo, no todos los modelos de fusión de sensores han demostrado ser capaces de lidiar con fallos en los sensores. De hecho, se ha demostrado que sistemas de fusión multi-sensor a nivel de características dejan de funcionar cuando falla un sensor. La fusión multi-sensor a nivel de decisiones se ha utilizado en trabajos relacionados para hacer frente a roturas de los sensores ya que en estos modelos se definen sistemas de reconocimiento independientes para cada sensor. No obstante, se ha demostrado en esta tesis que la decisión jerárquica y la decisión por mayoría, dos de los métodos más utilizados para la fusión a nivel de decisiones, presentan importantes limitaciones para superar los efectos de los fallos de los sensores. A partir de la identificación de las deficiencias de los modelos de decisión jerárquica y de decisión por mayoría se ha definido en este tesis una metodología novedosa que combina sus principales ventajas y evita sus principales limitaciones. Este modelo innovador denominado clasificador jerárquico ponderado (HWC) se ha evaluado en circunstancias ideales y se ha probado que proporciona un rendimiento comparable a los sistemas estándar de reconocimiento de la actividad. Incluso más importante que esto se ha demostrado que el HWC es capaz de paliar los efectos de los fallos de los sensores aún en el caso en que la mayoría de estos dejen de funcionar. Para el caso de la segunda categoría de las anomalías tecnológicas, esto es, defectos de los sensores, se ha evaluado la tolerancia del HWC y los modelos estándar de reconocimiento de la actividad a los efectos del cambio del rango dinámico de los sensores. Se ha demostrado que los sistemas basados en un solo sensor y los basados en la fusión multi-sensor a nivel de características sufren una degradación en su capacidad de reconocimiento cuando actúan en condiciones muy adversas. Además, se ha probado que los sistemas estándar de fusión multi-sensor a nivel de decisiones proporcionan una mayor tolerancia a los defectos en los sensores pero están aún lejos de ser capaces de proporcionar un reconocimiento preciso de la actividad. Por el contrario, se ha demostrado que el modelo HWC propuesto en esta tesis es tolerante a los defectos de los sensores cuando una minoría están dañados. De este modo, este modelo se demuestra útil para el reconocimiento de la actividad en configuraciones con sensores defectuosos.

El primer objetivo de esta tesis se ha cumplido con éxito ya que se ha evaluado el comportamiento de los sistemas estándar de reconocimiento de la actividad bajo los efectos de posibles anomalías tecnológicas de los sensores, y se ha contribuido con una solución alternativa e inno-

vadora para lidiar con los efectos de fallos y defectos imprevistos de los sensores.

Objetivo 2: Investigar la robustez de los sistemas estándar de reconocimiento de la actividad ante posibles variaciones imprevistas en el despliegue de los sensores, así como contribuir con una solución alternativa para lidiar con estas anomalías prácticas.

La mayoría de los sistemas de reconocimiento de la actividad asumen un despliegue predefinido de los sensores así como que este se mantiene sin cambios durante el uso del sistema. Sin embargo, estas suposiciones no son realistas puesto que los sensores pueden sufrir cambios en su distribución o emplazamiento en tanto en cuanto el usuario los coloca de la forma que considera más oportuna. Asimismo, los sensores pueden ser objeto de modificaciones con respecto a su emplazamiento original al estar integrados en ropa holgada o accesorios que no están firmemente sujetos al cuerpo. En estas circunstancias existe una alta probabilidad de que los sistemas de reconocimiento entrenados para un despliegue ideal de los sensores dejen de funcionar correctamente debido al desplazamiento de los sensores con respecto a su ubicación por defecto.

En este trabajo se ha introducido inicialmente el concepto de desplazamiento del sensor, el cual ha sido definido particularmente como una combinación de rotaciones y traslaciones. Además, se han presentado los principales efectos del desplazamiento en sensores portables inerciales. Se han planteado dos enfoques para el estudio de los efectos del desplazamiento de los sensores en los sistemas de reconocimiento de la actividad. En primer lugar, dicho efecto ha sido analizado para rotaciones y traslaciones modeladas de forma sintética. A partir de este análisis se ha demostrado que los sistemas de reconocimiento basados en un solo sensor son muy sensibles tanto a rotaciones como a traslaciones, especialmente aquellos que están concebidos para ser emplazados en partes del cuerpo sujetas a una reducida movilidad. Las configuraciones multi-sensor han demostrado una mayor tolerancia a los efectos del desplazamiento. Los modelos basados en fusión de características han demostrado una tolerancia moderada sólo para desplazamientos sutiles de los sensores, resultando de poca utilidad cuando más de un sensor es desplazado. A diferencia de estos, el modelo HWC propuesto en esta tesis ha demostrado ser capaz de afrontar variaciones moderadas en los sensores con independencia del número de sensores desplazados, así como proporcionar una alta tolerancia a grandes des-

plazamientos con respecto a su ubicación original cuando una minoría de los sensores están afectados. En segundo lugar, los efectos del desplazamiento de los sensores han sido investigados a partir de su observación en escenarios reales. Como parte de este trabajo, un nuevo conjunto o base de datos ha sido registrado para su uso en este tipo de evaluación dado que no existe ninguno disponible para el estudio del problema del desplazamiento de los sensores. La colección de esta base de datos ha sido concretamente ideada para explorar tanto los efectos de los desplazamientos de sensores introducidos por las personas al colocarse los dispositivos en el cuerpo, como el impacto de grandes desplazamientos que representan condiciones extremas para los algoritmos de reconocimiento. Para este conjunto de datos se ha realizado un análisis estadístico de los efectos del desplazamiento mostrando de forma cuantitativa las variaciones que dichos desplazamientos introducen en las señales con respecto a un despliegue ideal de los sensores. Este estudio ha permitido comprobar la marcada disparidad en las distribuciones de los datos cuando los usuarios ubican los sensores de forma arbitraria, la cual se acrecienta de forma substancial cuando los sensores son ampliamente desplazados con respecto a su ubicación predeterminada. Dichos efectos también han sido analizados desde el punto de vista del rendimiento de los sistemas de reconocimiento de la actividad en presencia de desplazamiento de los sensores. Se ha comprobado que ni los sistemas de reconocimiento basados en un sensor ni aquellos que se basan en la fusión de múltiples sensores a nivel de características son capaces de lidiar con los efectos del desplazamiento de los sensores. A diferencia de los anteriores, el HWC sí ha demostrado una capacidad destacada para asimilar los cambios introducidos por la colocación de los sensores por parte de los usuarios, así como para operar de forma moderada en aquellas situaciones en las que los sensores son significativamente desplazados.

El segundo objetivo de esta tesis ha sido alcanzado puesto que se ha realizado la evaluación del comportamiento de sistemas estándar de reconocimiento de la actividad ante los efectos de cambios en el despliegue de los sensores y se ha planteado una posible alternativa para lidiar con las variaciones introducidas por el desplazamiento de los sensores.

Objetivo 3: Estudiar la capacidad de los sistemas estándar de reconocimiento de la actividad para soportar posibles cambios en la red de sensores, así como contribuir con una solución alternativa para lidiar con estas variaciones topológicas.

En el mundo real, los sistemas de reconocimiento de la actividad están sujetos a variaciones en el equipamiento. Por una parte, sensores obsoletos o dañados pueden ser reemplazados por otros de características diferentes durante operaciones de mantenimiento. Igualmente, nuevos sensores pueden ser incorporados a la red de sensores original como parte de procesos de actualización del sistema. Finalmente, la disponibilidad de los sensores puede variar durante el curso normal de día dependiendo del contexto particular en el que se encuentra la persona. En estos escenarios, los nuevos sensores incorporados a la topología de sensores concebida durante la fase de diseño carecen en principio de capacidades de reconocimiento de la actividad puesto que no tienen asociado un sistema para ello. En consecuencia, es necesaria la generación de nuevos sistemas de reconocimiento de las actividades de interés a través de un entrenamiento específico para los nuevos sensores. Este entrenamiento requiere a su vez del registro de bases de datos específicas, lo cual resulta inabordable en escenarios reales.

En este trabajo se han analizado las principales limitaciones de las técnicas de transferencia de conocimiento planteadas en la literatura como solución al entrenamiento de nuevos sistemas de reconocimiento. Teniendo en cuenta estas limitaciones (alta latencia, falta de generalización, unimodalidad), se ha propuesto un novedoso método que permite transferir de forma automática las capacidades de un sistema de reconocimiento existente o fuente, a un sistema de reconocimiento no entrenado u objetivo, incluso cuando las modalidades de sensado de fuente y objetivo difieren. El modelo propuesto se basa en el aprendizaje de una función de mapeado entre las señales del sensor fuente y el sensor objetivo, la cual es utilizada para transferir modelos de reconocimiento de actividad o patrones de reconocimiento de actividad dependiendo del modo de operación del sistema de reconocimiento fuente. El método propuesto ha sido evaluado en un escenario multimodal para el reconocimiento de gestos, en el que se utilizan tanto sensores ambientales como vestibles. Se ha estudiado tanto la transferencia de capacidades de reconocimiento entre sensores de la misma modalidad como entre sensores de distinta modalidad. Para el primer caso, la transferencia entre sensores vestibles ha sido evaluada, demostrándose el éxito del modelo tanto para la transferencia de modelos de reconocimiento o patrones de reconocimiento entre sensores ubicados en miembros adyacentes. Igualmente, se ha comprobado que la transferencia de modelos y patrones entre sistemas de reconocimiento operando en sensores ambientales y vestibles es satisfactoria. Tanto para la transferencia entre sensores de

la misma modalidad como sensores de distinta modalidad se ha demostrado que la información proporcionada por un solo gesto es suficiente para que el modelo de transferencia aprenda la relación física que permita mapear las señales fuente y objetivo. Además, este mapeado es generalizable a otros movimientos o gestos diferentes al realizado para el aprendizaje de dicha relación. Se ha comprobado asimismo que se necesita de una mayor cantidad de información para el aprendizaje de la función de mapeado cuando el usuario se comporta de forma libre o arbitraria. No obstante, la cantidad de información necesaria en dicha situación es muy inferior a la requerida por otros modelos propuestos en trabajos previos. Además de permitir una transferencia exitosa de las capacidades de reconocimiento, esta se realiza de forma rápida y sin requerir la intervención del usuario o el diseñador del sistema en ningún caso. Estas características son especialmente interesantes atendiendo a las necesidades de los sistemas de reconocimiento del mundo real.

El tercer objetivo de esta tesis ha sido completado puesto que las principales limitaciones de los modelos clásicos de entrenamiento de nuevos sistemas de reconocimiento han sido identificadas, y un nuevo modelo ha sido propuesto para suplir dichas limitaciones proporcionando un medio para el aprendizaje de nuevos sistemas de reconocimiento en condiciones reales.

7.2. Contribuciones

En la Sección 7.1 se ha descrito que los objetivos de esta tesis se han cumplido de forma satisfactoria. Las principales contribuciones de esta tesis se listan a continuación:

- Identificación de los requerimientos y retos planteados por los sistemas de reconocimiento de la actividad en condiciones reales.
- Evaluación de la tolerancia de los sistemas de reconocimiento estándar a los efectos de anomalías tecnológicas de los sensores, particularmente fallos y defectos.
- Definición y desarrollo de un modelo novedoso (HWC) para lidiar con los efectos de fallos y defectos en los sensores.
- Evaluación de la robustez del modelo HWC a los efectos de fallos y defectos en los sensores.

- Evaluación de la tolerancia de los sistemas de reconocimiento estándar a variaciones en el despliegue de los sensores, particularmente desplazamientos estáticos y dinámicos.

- Evaluación de la robustez del modelo HWC a los efectos de desplazamientos de los sensores.

- Definición, desarrollo y validación de un sistema novedoso de transferencia de conocimiento multimodal, capaz de operar en tiempo real, con poca carga computacional y sin necesidad de intervención alguna por parte del usuario o diseñador del sistema.

- Registro y adecuación de un conjunto de datos destinados a investigar los efectos de variaciones en el despliegue de los sensores (desplazamiento de los sensores). Esta base de datos contiene información correspondiente a un amplio rango de actividades, sensores y sujetos. Además de su uso para la investigación del problema del desplazamiento de los sensores, la base de datos registrada en este trabajo se puede utilizar también para comparar técnicas de reconocimiento de la actividad en condiciones ideales. La base de datos está disponible en http://www.ugr.es/~oresti/datasets.

- Registro y adecuación de un conjunto de datos destinados a investigar la transferencia de conocimiento multimodal entre sistemas de reconocimiento basados en sensores ambientales y sensores vestibles. La base de datos puede ser también utilizada para detección de gestos y reconocimiento de la actividad. La base de datos está disponible en http://www.ugr.es/~oresti/datasets.

7.3. Trabajo futuro

Teniendo en cuenta la novedad de este trabajo existe todavía un amplio margen para la investigación de nuevos modelos de reconocimiento de la actividad para el mundo real. En esta sección se presentan diversas líneas de trabajo futuro identificadas como continuación o extensión de la investigacin desarrollada en esta tesis.

7.3.1. Registro de nuevas bases de datos para validación de modelos de reconocimiento

Una de las principales limitaciones en el campo del reconocimiento de la actividad humana basada en sensores vestibles es la falta de bases de datos que permitan comparar los nuevos modelos propuestos con los métodos desarrollados en trabajos previos. En consecuencia, la comunidad científica debe hacer un esfuerzo importante para tratar de generar nuevas bases de datos que puedan ser libremente utilizadas para validar soluciones pasadas, presentes y futuras al problema del reconocimiento de la actividad humana. De forma más concreta, se debería plantear la creación de bases de datos que reflejen los problemas que deben afrontar los sistemas de reconocimiento en el mundo real. Por ejemplo, bases de datos que incluyan anomalías de tipo tecnológico como fallos o defectos son especialmente valiosas para testear modelos diseñados para lidiar con los efectos de dichas anomalías. Asimismo, sería de una gran utilidad una extensión de la base de datos proporcionada en este trabajo para investigar los efectos del desplazamiento de los sensores. Concretamente, resultaría de especial interés observar el desplazamiento introducido durante el uso habitual de los artículos cotidianos que incluyen estos sensores. Finalmente, el uso de protocolos de monitorización masiva, esto es, cientos o miles de personas, se considera de especial trascendencia de cara a obtener una gran cantidad de información que permita generalizar la validación de los modelos propuestos.

7.3.2. Reconfiguración dinámica del HWC

Son muchas las propiedades a destacar del modelo HWC propuesto en esta tesis. De entre estas, una de las más importantes es su flexibilidad a variaciones en el equipamiento de sensado. Este modelo ha demostrado que los sensores pueden ser eliminados o añadidos manteniendo la capacidad de reconocimiento en todo momento. Esto es especialmente importante de cara a soportar la integración de futuras tecnologías de sensado no conocidas durante la fase de diseño del sistema.

Los sistemas de reconocimiento son normalmente diseñados para la detección e identificación de un conjunto de actividades específicas. No obstante, las actividades de interés pueden cambiar durante el curso del tiempo dependiendo de las necesidades particulares del usuario. Por ejemplo, actividades adicionales a las consideradas durante la fase de diseño pueden ser requeridas cuando los ejercicios de un plan de en-

trenamiento son modificados. Estos cambios pueden conllevar no sólo la inclusión de nuevas actividades sino también la eliminación de otras que pasan a no ser objetivo específico del usuario. Esto es de especial interés de cara a reducir la complejidad de los sistemas de reconocimiento e incrementar su eficiencia, así como para permitir la personalización de dichos sistemas a los intereses y necesidades de cada usuario. Este tipo de cambios no son aplicables en los modelos de reconocimiento estándar puesto que cualquier variación en el conjunto de actividades objetivo implica un reentrenamiento completo del sistema. A diferencia de estos, la flexibilidad del modelo HWC tolera modificaciones de este tipo. De hecho, sólo es necesario entrenar el clasificador base asociado a la nueva actividad y actualizar los pesos en concordancia. En el caso de que una actividad sea eliminada, el proceso simplemente consiste en quitar el clasificador asociado a esta y recalcular los pesos. Parte del trabajo futuro de esta tesis tiene como objetivo la evaluación de la capacidad del modelo HWC en esta dirección.

7.3.3. HWC auto-adaptativo

El modelo HWC propuesto en esta tesis ha demostrado una importante tolerancia a los errores introducidos por anomalías en los sensores. No obstante, este método tiene en cuenta las decisiones de todos los sensores considerados durante el proceso de diseño, de modo que su robustez se ve afectada cuando una mayoría de sensores son modificados de forma sustancial. La definición de un modelo adaptativo que permita actualizar de forma autónoma los pesos asociados a cada uno de los sensores permitiría reducir el impacto de aquellas decisiones adoptadas sobre la información registrada por sensores anómalos. Sin embargo, dicho proceso de adaptación requiere de un conocimiento extra o realimentación para determinar qué sensores se comportan de forma anómala. Dicha información podría ser proporcionada directamente por el usuario, el cual puede ser consciente de si el sistema está acertando en el proceso de reconocimiento o no. No obstante, esto representa una tarea tediosa que puede suponer una verdadera molestia para el usuario. Por otra parte, se podría plantear la utilización del conocimiento obtenido de la colectividad de sensores para determinar cuál o cuáles se comportan de forma indeseada. Este tipo de técnicas permitirían la auto-adaptación del modelo HWC a las condiciones concretas del equipamiento utilizado.

Reducir los pesos para aquellos sensores identificados como anómalos

o incluso eliminarlos del proceso de toma de decisiones puede ser una buena práctica para mantener la capacidad de reconocimiento. Sin embargo, las anomalías que experimentan los sensores no siempre afectan de la misma forma al proceso de reconocimiento de las diversas actividades consideradas. De hecho, algunas actividades pueden verse más afectadas por fallos en los sensores que otras tal y como se ha mostrado en esta tesis. Asimismo, el desplazamiento de un sensor con respecto a su emplazamiento predeterminado puede hacer imposible la identificación de algunas actividades mientras que otras pueden ser aun perfectamente reconocibles. Teniendo en cuenta la flexibilidad de la estructura del HWC sería posible reducir el valor de los pesos asociados a aquellas actividades que se ven alteradas por las citadas anomalías, manteniendo los pesos para aquellas otras que se siguen reconociendo sin ningún tipo de problema. Este tipo de procedimiento es bastante más preciso que la reducción de los pesos a nivel de sensor o incluso su eliminación en la toma de decisiones, y permite aprovechar el potencial de los sensores afectados para las restantes actividades reconocibles. Parte del trabajo futuro de esta tesis tiene como objetivo la definición de un mecanismo de auto-adaptación que extienda las capacidades del modelo HWC en esta dirección.

7.3.4. Tolerancia a otras anomalías tecnológicas y topológicas

Además de las anomalías investigadas en esta tesis existen otras que también sería importante estudiar. Por ejemplo, algunas anomalías de tipo tecnológico son más dependientes de cómo se registran y procesan las señales de forma simultánea en las configuraciones multi-sensor. Aunque normalmente se asume que las señales de movimiento son registradas por cada sensor a una tasa de muestreo similar, puede suceder que exista desincronía entre las señales al ser referidas a una misma referencia temporal. Para evitar este tipo de desviaciones temporales normalmente se utilizan configuraciones cableadas de los sensores, lo cual es totalmente impráctico para contextos reales. En las redes inalámbricas se usa una misma referencia temporal para cada sensor sin embargo normalmente aparecen desviaciones entre estas. La evaluación de estas derivas temporales, especialmente en despliegues con multitud de sensores, representa un interesante reto a considerar como trabajo futuro de esta tesis.

Los efectos del desplazamiento de los sensores han sido particularmente analizados para datos de aceleración. La investigación de dichos

efectos en otras modalidades inerciales resulta igualmente interesante. De hecho, en un trabajo reciente [144] se ha proporcionado una evaluación preliminar de los efectos de desplazamiento para diversas modalidades inerciales. Además de esto, como trabajo futuro de esta tesis se plantea el análisis de los efectos de desplazamiento extremo de los sensores, esto es, cuando dos o más sensores son intercambiados de sus posiciones originales de una forma drástica. Este sería el caso en el que un usuario coloca un brazalete idealmente diseñado para ser puesto en la muñeca en otra parte del cuerpo, por ejemplo el tobillo. Este tipo de investigación se prevé de especial importancia de cara a dar la máxima libertad al usuario a la hora de utilizar los artículos que incluyen los sensores de medición del movimiento.

7.3.5. Transferencia de conocimiento entre múltiples modalidades de sensado

En esta tesis se ha analizado el mapeo entre sensores vestibles y entre sensores vestibles y ambientales. Concretamente se han caracterizado modelos de mapeo de aceleración a aceleración y aceleración a posición. Si bien este es un problema complejo *per se*, otras combinaciones más complejas son posibles. Por ejemplo, sensores ambientales como los utilizados para detectar la interacción del usuario con el entorno (por ejemplo, pulsadores) son habitualmente utilizados para inferir actividades de tipo doméstico. El mapeo desde esta modalidad ambiental a una modalidad de sensor de tipo vestible puede resultar inconcebible. No obstante, el uso de modelos de mapeo más complejos podría permitir encontrar la relación entre este tipo de dominios complejos. Este tipo de modelos podría beneficiarse de la combinación de fuentes de información heterogéneas que permitan describir mejor el fenómeno observado. Esto permitiría la definición de funciones de mapeo más precisas, si bien ello se lograría a expensas de un modelado más complejo.

El modelado de las transformaciones entre múltiples sensores (por ejemplo, de dos medidas de posición a una de aceleración) es otra línea de investigación futura interesante. En esta tesis se ha demostrado la importancia de monitorizar diferentes partes del cuerpo de cara a proporcionar un reconocimiento preciso y robusto de las actividades ejecutadas por el individuo. De forma análoga, el uso de múltiples fuentes de información o sensores se plantea aquí para mejorar la capacidad de generalización y robustez de la función de mapeado.

Para poder hacer uso de múltiples fuentes de información o sensores

resulta necesario utilizar modelos más complejos a los presentados en esta tesis. La selección dinámica del mejor conjunto de señales de entrada para obtener un mapeo óptimo también puede ayudar a conseguir unos mejores resultados. En cualquier caso, la definición de modelos sofisticados como los presentados en esta tesis abren un nuevo espectro de posibilidades en la investigación de sistemas de reconocimiento de la actividad para el mundo real.

Glossary

Activity Recognition Chain set of tools used for the AR process.

Black-box device, system or object which can be viewed in terms of its input, output and transfer characteristics without any knowledge of its internal workings.

Body Sensor Network also known as a wireless body area network (WBAN) or body area network (BAN), refers to a wireless network that consists of several wearable computing devices..

Gray-box device, system or object for which inputs, outputs and internal features are partially-known. To construct the model GB combines both (system insights) and (experimental data) techniques.

Human-Computer Interaction comprises the study, planning, and design of the interaction between people (users) and computers (machines).

White-box device, system or object for which inputs, outputs and internal features are well-known and can be modeled.

Accelerometer device that measures proper acceleration. MEMS accelerometers are increasingly present in portable electronic devices to detect the position and orientation of the device, as well as to register the motion it is subject to..

Gyroscope device that measures proper orientation based on the principles of angular momentum and rate of turn. MEMS gyroscopes takes the idea of the Foucault pendulum and uses a vibrating element. They can be found on some consumer electronic devices and are normally included in inertial navigation systems..

Inertial Measurement Unit device that measures the proper velocity, orientation, and gravitational forces, using a combination of accelerometers and gyroscopes, sometimes also magnetometers. IMUs are specially utilized in aircraft, spacecraft, watercraft, and

guided missiles, and during the last years also part of on-body AR systems..

Magnetometer device that measures the direction of the magnetic field at a point in space. Although magnetometers are utilized in a wide range of applications, in AR these systems serve as a compass. They can be found on some consumer electronic devices, principally in smartphones..

Microelectromechanical systems also referred to as micro systems technology, is the technology of very small devices. It normally consists of a central unit that processes data and several components that interact with the surroundings such as microsensors. MEMS devices are highly exploited in markets such as automobiles, biomedical, and electronics, being examples of these systems accelerometers, gyroscopes, magnetometers or microphones among others..

Acronyms

ANN Artificial Neural Network.

AR (Human) Activity Recognition.

ARC Activity Recognition Chain.

ARMA Autoregressive-moving-average.

ARMAX Autoregressive-moving-average with exogenous inputs.

BB Black-box.

BSN Body Sensor Network.

DFMARC Decision Fusion Multi-Sensor Activity Recognition Chain.

DT Decision Trees.

FFMARC Feature Fusion Multi-Sensor Activity Recognition Chain.

GB Gray-box.

HCI Human-Computer Interaction.

HD Hierarchical Decision.

HWC Hierarchical Weighted Classifier.

IMU Inertial Measurement Unit.

KNN K-Nearest Neighbor.

MARC Multiple Sensor Activity Recognition Chain.

MEMS Microelectromechanical systems.

MIMO Multiple-input-multiple-output.

MV Majority Voting.

NB Naive Bayes.

SARC Single Sensor Activity Recognition Chain.

SVM Support Vector Machines.

WB White-box.

List of Figures

List of Tables

Bibliography

[1] T. E. Starner. The cyborgs are coming, or, the real personal computers. Technical report, Massachusets Institute of Technology, 1994.

[2] T. E. Starner. Project Glass: An extension of the self. *IEEE Pervasive Computing*, 12(2):14–16, 2013.

[3] D. Roggen, S. Magnenat, M. Waibel, and G. Tröster. Wearable computing: Designing and sharing activity-recognition systems across platforms. *IEEE Robotics and Automation Magazine*, 18(2), 2011.

[4] W. Barfield and T. Caudell, editors. *Fundamentals of Wearable Computers and Augmented Reality*. CRC Press, 2001.

[5] D. Figo, P. C. Diniz, D. R. Ferreira, and J. M. P. Cardoso. Preprocessing techniques for context recognition from accelerometer data. *Personal and Ubiquitous Computing*, 14(7):645–662, 2010.

[6] W. Wang, Y. W. Guo, B. Y. Huang, G. R. Zhao, B. Q. Liu, and L. Wang. Analysis of filtering methods for 3D acceleration signals in body sensor network. In *IEEE International Symposium on Bioelectronics and Bioinformatics*, pages 263–266, 2011.

[7] S. Pirttikangas, K. Fujinami, and T. Seppanen. Feature selection and activity recognition from wearable sensors. In *Third International Symposium Ubiquitous Computing Systems*, volume 4239 of *LNCS*, pages 516–527, 2006.

[8] A. Mannini, S. S. Intille, M. Rosenberger, A. M. Sabatini, and W. Haskell. Activity recognition using a single accelerometer placed at the wrist or ankle. *Medicine and Science in Sports and Exercise*, 45(11):2193–2203, 2013.

[9] M. Stikic, T. Huynh, K. Van Laerhoven, and B. Schiele. ADL recognition based on the combination of RFID and accelerometer sensing. In *Proceedings of the 2nd International Conference on Pervasive Computing Technologies for Healthcare*, page 258–263, Tampere, Finland, January 2008.

[10] L. Bao and S. S. Intille. Activity recognition from user-annotated acceleration data. In *Pervasive Computing*, volume 23, pages 1–17, 2004.

[11] R. Marx. Ad-hoc accelerometer activity recognition in the iball. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, 2010.

[12] Z. He, Z. Liu, J. Lianwen, L.-X. Zhen, and J.-C. Huang. Weightlessness feature - a novel feature for single tri-axial accelerometer based activity recognition. In *19th International Conference on Pattern Recognition*, pages 1–4, 2008.

[13] M. J. Mathie, A. C. F. Coster, N. H. Lovell, and B. G. Celler. Detection of daily physical activities using a triaxial accelerometer. *Medical and Biological Engineering and Computing*, 41(3):296–301, 2003.

[14] J. Baek, G. Lee, W. Park, and B. Y. Yun. Accelerometer signal processing for user activity detection. *Knowledge-Based Intelligent Information and Engineering Systems*, pages 610–617, 2004.

[15] U. Maurer, A. Smailagic, D. P. Siewiorek, and M. Deisher. Activity recognition and monitoring using multiple sensors on different body positions. In *International Workshop on Wearable and Implantable Body Sensor Networks*, pages 113–116, April 2006.

[16] N. Ravi, P. Mysore, and M. L. Littman. Activity recognition from accelerometer data. In *Proceedings of the 17th Conference on Innovative Applications of Artificial Intelligence*, pages 1541–1546, 2005.

[17] M. N. Nyan, F. E. H. Tay, K. H. W. Seah, and Y. Y. Sitoh. Classification of gait patterns in the timefrequency domain. *Journal of Biomechanics*, 39(14):2647–2656, 2006.

[18] S. J. Preece, J. Y. Goulermas, L. P. J. Kenney, and D. Howard. A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. *IEEE Transactions on Biomedical Engineering*, 56(3):871–879, march 2009.

[19] J. Mäntyjärvi, J. Himberg, and T. Seppanen. Recognizing human motion with multiple acceleration sensors. In *IEEE International*

Conference on Systems, Man and Cybernetics, pages 747–752, 2001.

[20] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182, March 2003.

[21] H. K. Lee and J. H. Kim. An HMM-based threshold model approach for gesture recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(10):961–973, October 1999.

[22] P. Lukowicz, J. A. Ward, H. Junker, M. Stäger, G. Tröster, A. Atrash, and T. Starner. Recognizing workshop activity using body worn microphones and accelerometers. *In Pervasive Computing*, pages 18–32, April 2004.

[23] P. Boissy, S. Choquette, M. Hamel, and N. Noury. User-based motion sensing and fuzzy logic for automated fall detection in older adults. *Telemedicine and e-Health*, 13(6):683–694, 2007.

[24] J. Lester, T. Choudhury, N. Kern, G. Borriello, and B. Hannaford. A hybrid discriminative/generative approach for modeling human activities. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence*, pages 766–772, Edinburgh, Scotland, 2005.

[25] A. Reiss, G. Hendeby, and D. Stricker. Confidence-based multiclass adaboost for physical activity monitoring. In *Proceedings of the 2013 International Symposium on Wearable Computers*, pages 13–20. ACM, 2013.

[26] J. Parkka, M. Ermes, P. Korpipaa, J. Mantyjarvi, J. Peltola, and I. Korhonen. Activity classification using realistic data from wearable sensors. *IEEE Transactions on Information Technology in Biomedicine*, 10(1):119–128, 2006.

[27] M. Ermes, J. Parkka, J. Mantyjarvi, and I. Korhonen. Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. *IEEE Transactions on Information Technology in Biomedicine*, 12(1):20–26, 2008.

[28] K. Förster, S. Monteleone, A. Calatroni, D. Roggen, and G. Tröster. Incremental knn classifier exploiting correct - error teacher for activity recognition. In *Procedings of the 9th International Conference on Machine Learning and Applications*, pages 445–450, 2010.

[29] J. Lester, T. Choudhury, N. Kern, G. Borriello, and B. Hannaford. A hybrid discriminative/generative approach for modeling human activities. In *Proceedings of the 19th international joint conference on Artificial intelligence*, IJCAI'05, pages 766–772, 2005.

[30] J. Parera, C. Angulo, A. Rodriguez-Molinero, and J. Cabestany. User daily activity classification from accelerometry using feature selection and svm. In *Bio-Inspired Systems: Computational and Ambient Intelligence*, LNCS, pages 1137–1144, 2009.

[31] H. He and E. A. Garcia. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9):1263–1284, 2009.

[32] S. Arlot and A. Celisse. A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4:40–79, 2010.

[33] L. Breiman and P. Spector. Submodel selection and evaluation in regression. the x-random case. *International Statistical Review*, 60(3):291–319, 1992.

[34] R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pages 1137–1143, 1995.

[35] M. Stone. Asymptotics for and against cross-validation. *Biometrika*, pages 29–35, 1977.

[36] M. Sokolova and G. Lapalme. A systematic analysis of performance measures for classification tasks. *Information Processing and Management*, 45(4):427 – 437, 2009.

[37] A. M. Khan, Y. K. Lee, S. Y. Lee, and T. S. Kim. A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer. *IEEE Transactions on Information Technology in Biomedicine*, 14(5):1166–1172, September 2010.

[38] A. Akl, C. Feng, and S. Valaee. A novel accelerometer-based gesture recognition system. *IEEE Transactions on Signal Processing*, 59(12):6197–6205, December 2011.

[39] J. R. Kwapisz, G. M. Weiss, and S. A. Moore. Activity recognition using cell phone accelerometers. *17th Conference on Knowledge Discovery and Data Mining*, 12(2):74–82, 2011.

[40] Y. Fu, L. Cao, G. Guo, and T. S. Huang. Multiple feature fusion by subspace learning. In *Proceedings of the 2008 International Conference on Content-based Image and Video Retrieval*, pages 127–134, 2008.

[41] K. Van Laerhoven, A. Schmidt, and H.-W. Gellersen. Multi-sensor context aware clothing. In *Proc. of the 6th International Symposium on Wearable Computers*, pages 49–56, 2002.

[42] P. Lukowicz, H. Junker, M. Staeger, T. von Bueren, and G. Troester. WearNET: A distributed multi-sensor system for context aware wearables. In *Proceedings of the 4th International Conference on Ubiquitous Computing*, pages 361–370, September 2002.

[43] N. Kern, B. Schiele, and A. Schmidt. Multi-sensor activity context detection for wearable computing. In *European Symposium on Ambient Intelligence*, pages 220–232, Eindhoven, The Netherlands, November 2003.

[44] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):226–239, March 1998.

[45] D. Ruta and B. Gabrys. An overview of classifier fusion methods. *Computing and Information Systems*, 7:1–10, 2000.

[46] J. Fahrenberg, F. Foerster, M. Smeja, and W. Miller. Assessment of posture and motion by multichannel piezoresistive accelerometer recordings. *Psychophysiology*, 34(5):607–612, 1997.

[47] S. H. Lee, H. D. Park, S. Y. Hong, K. J. Lee, and Y. H. Kim. A study on the activity classification using a triaxial accelerometer. In *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 3, pages 2941–2943 Vol.3, September 2003.

[48] M. J. Mathie, A. C. F. Coster, N. H. Lovell, B. G. Celler, S. R. Lord, and A. Tiedemann. A pilot study of long-term monitoring of human movements in the home using accelerometry. *Journal of telemedicine and telecare*, 10(3):144–151, 2004.

[49] D. M. Karantonis, M. R. Narayanan, M. Mathie, N. H. Lovell, and B. G. Celler. Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring. *IEEE Transactions on Information Technology in Biomedicine*, 10(1):156–167, 2006.

[50] L. Gao, A. K. Bourke, and J. Nelson. Activity recognition using dynamic multiple sensor fusion in body sensor networks. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1077–1080, September 2012.

[51] P. Zappi, T. Stiefmeier, E. Farella, D. Roggen, L. Benini, and G. Troster. Activity recognition from on-body sensors by classifier fusion: sensor scalability and robustness. In *3rd International Conference on Intelligent Sensors, Sensor Networks and Information*, pages 281–286, December 2007.

[52] R. Chavarriaga, H. Sagha, and J. del R. Millan. Ensemble creation and reconfiguration for activity recognition: An information theoretic approach. In *IEEE International Conference on Systems, Man, and Cybernetics*, pages 2761–2766, October 2011.

[53] H. Sagha, J. del R. Millan, and R. Chavarriaga. Detecting and rectifying anomalies in opportunistic sensor networks. In *8th Int. Conf. on Networked Sensing Systems*, pages 162–167, Penghu, Taiwan, June 2011.

[54] P. Zappi, D. Roggen, E. Farella, G. Tröster, and L. Benini. Network-level power-performance trade-off in wearable activity recognition: A dynamic sensor selection approach. *ACM Transactions on Embedded Computer Systems*, 11(3):68:1–68:30, September 2012.

[55] H. Sagha, H. Bayati, J. del R. Millan, and R. Chavarriaga. On-line anomaly detection and resilience in classifier ensembles. *Pattern Recognition Letters*, 34(15):1916–1927, 2013.

[56] H. Sagha, A. Calatroni, J. del R. Millan, D. Roggen, G. Troster, and R. Chavarriaga. Robust activity recognition combining anomaly detection and classifier retraining. In *IEEE International Conference on Body Sensor Networks*, pages 1–6, 2013.

[57] K. Ni, N. Ramanathan, M. N. H. Chehade, L. Balzano, S. Nair, S. Zahedi, E. Kohler, G. Pottie, M. Hansen, and M. Srivastava. Sensor network data fault types. *ACM Transactions on Sensor Networks*, 5(3):25:1–25:29, June 2009.

[58] N. Ramanathan, K. Chang, R. Kapur, L. Girod, E. Kohler, and D. Estrin. Sympathy for the sensor network debugger. In *3rd International Conference on Embedded Networked Sensor Systems*, SenSys '05, pages 255–267, 2005.

[59] S. Rost and H. Balakrishnan. Memento: A health monitoring system for wireless sensor networks. In *3rd Annual IEEE Communications Society on Sensor and Ad Hoc Communications and Networks*, volume 2, pages 575–584, 2006.

[60] N. Ramanathan, T. Schoellhammer, E. Kohler, K. Whitehouse, T. Harmon, and D. Estrin. Suelo: human-assisted sensing for exploratory soil monitoring studies. In *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems*, pages 197–210, 2009.

[61] Y. Yao, A. Sharma, L. Golubchik, and R. Govindan. Online anomaly detection for sensor systems: A simple and efficient approach. *Performance Evaluation*, 67(11):1059–1075, November 2010.

[62] J. Chen, S. Kher, and A. Somani. Distributed fault detection of wireless sensor networks. In *Proceedings of the 2006 Workshop on Dependability Issues in Wireless Ad Hoc Networks and Sensor Networks*, pages 65–72, 2006.

[63] S. Rajasegarar, C. Leckie, M. Palaniswami, and J.C. Bezdek. Distributed anomaly detection in wireless sensor networks. In *10th IEEE Singapore International Conference on Communication systems*, pages 1–5, October 2006.

[64] S. Rajasegarar, C. Leckie, M. Palaniswami, and J. C. Bezdek. Quarter sphere based distributed anomaly detection in wireless

sensor networks. In *IEEE International Conference on Communications*, pages 3864–3869, June 2007.

[65] S. Ganeriwal, L. K. Balzano, and M. B. Srivastava. Reputation-based framework for high integrity sensor networks. *ACM Transaction on Sensor Networks*, 4(3):1–37, June 2008.

[66] F. Koushanfar, M. Potkonjak, and A. Sangiovanni-Vincentelli. On-line fault detection of sensor measurements. In *Proceedings of IEEE Sensors Conference*, volume 2, pages 974–979 Vol.2, October 2003.

[67] K. Murao, T. Terada, Y. Takegawa, and S. Nishio. A context-aware system that changes sensor combinations considering energy consumption. In Jadwiga Indulska, Donald J. Patterson, Tom Rodden, and Max Ott, editors, *Pervasive*, volume 5013 of *Lecture Notes in Computer Science*, pages 197–212. Springer, 2008.

[68] R. Uchida, H. Horino, and R. Ohmura. Improving fault tolerance of wearable wearable sensor-based activity recognition techniques. In *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication*, pages 633–644, 2013.

[69] T.-Y. Wang, Y. S. Han, P. K. Varshney, and P.-N. Chen. Distributed fault-tolerant classification in wireless sensor networks. *IEEE Journal on Selected Areas in Communications*, 23(4):724–734, 2005.

[70] K. Kapitanova, E. Hoque, J. A. Stankovic, K. Whitehouse, and S. H. Son. Being smart about failures: Assessing repairs in smart homes. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 51–60, 2012.

[71] L. Gao, A. K. Bourke, and J. Nelson. A system for activity recognition using multi-sensor fusion. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 7869–7872, August 2011.

[72] K. Kunze and P. Lukowicz. Dealing with sensor displacement in motion-based onbody activity recognition systems. In *10th international conference on Ubiquitous computing*, pages 20–29, Seoul, South Korea, September 2008.

[73] K. Förster, P. Brem, D. Roggen, and G. Troster. Evolving discriminative features robust to sensor displacement for activity recognition in body area sensor networks. In *5th International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, pages 43–48, December 2009.

[74] U. Steinhoff and B. Schiele. Dead reckoning from the pocket - an experimental study. In *IEEE International Conference on Pervasive Computing and Communications*, pages 162–170, March 2010.

[75] H. Bayati, J. del R. Millan, and R. Chavarriaga. Unsupervised adaptation to on-body sensor displacement in acceleration-based activity recognition. In *15th Annual International Symposium on Wearable Computers*, pages 71–78, June 2011.

[76] R. Chavarriaga, H. Bayati, and J. del R. Millan. Unsupervised adaptation for acceleration-based activity recognition: robustness to sensor displacement and rotation. *Pers Ubiquit Comput*, pages 1–12, 2012. Online first.

[77] M. Sugiyama, M. Krauledat, and K.-R. Müller. Covariate shift adaptation by importance weighted cross validation. *Journal of Machine Learning Research*, 8:985–1005, December 2007.

[78] K. Förster, D. Roggen, and G. Tröster. Unsupervised classifier self-calibration through repeated context occurences: Is there robustness against sensor displacement to gain? In *13th IEEE Int. Symposium on Wearable Computers*, pages 77–84, Linz, Austria, September 2009.

[79] P. Lukowicz, J. A. Ward, H. Junker, M. Stäger, G. Tröster, A. Atrash, and T. Starner. Recognizing workshop activity using body worn microphones and accelerometers. *Pervasive Computing*, pages 18–32, April 2004.

[80] P. Lukowicz, F. Hanser, C. Szubski, and W. Schobersberger. Detecting and interpreting muscle activity with wearable force sensors. In *Pervasive Computing*, pages 101–116, 2006.

[81] O. Amft. A wearable earpad sensor for chewing monitoring. In *Proc. of IEEE Sensors Conference*, pages 222–227, 2010.

[82] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. T. Campbell. A survey of mobile phone sensing. *IEEE Communications Magazine*, 48(9):140–150, 2010.

[83] G. Kortuem, F. Kawsar, D. Fitton, and V. Sundramoorthy. Smart objects as building blocks for the internet of things. *IEEE Internet Computing*, 14(1):44–51, 2010.

[84] H. Harms, O. Amft, and G. Tröster. Estimating posture recognition performance in sensing garments using geometric wrinkle modeling. *IEEE Transactions on Information Technology in Biomedicine*, 14(6):1436–1445, 2010.

[85] T. L. M. Van Kasteren, G. Englebienne, and B. J. A. Kröse. Transferring knowledge of activity recognition across sensor networks. In *Proc. 8th Int. Conf on Pervasive Computing*, pages 283–300, 2010.

[86] L. Xia, C.-C. Chen, and J. K. Aggarwal. Human detection using depth information by kinect. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 15–22, June 2011.

[87] J. Lester, T. Choudhury, and G. Borriello. A practical approach to recognizing physical activities. In *Proc. of Pervasive Computing*, pages 1–16, 2006.

[88] K. Kunze, G. Bahle, P. Lukowicz, and K. Partridge. Can magnetic field sensors replace gyroscopes in wearable sensing applications? In *Proc. 2010 Int. Symp. on Wearable Computers*, 2010.

[89] K. Kunze, P. Lukowicz, H. Junker, and G. Troester. Where am i: Recognizing on-body positions of wearable sensors. *International Workshop on Location and Context-Awareness*, January 2005.

[90] K. Kunze, P. Lukowicz, K. Partridge, and B. Begole. Which way am i facing: Inferring horizontal device orientation from an accelerometer signal. In *13th IEEE Int. Symposium on Wearable Computers*, pages 149–150, 2009.

[91] S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 99(1), 2010.

[92] A. Calatroni, C. Villalonga, D. Roggen, and G. Tröster. Context cells: Towards lifelong learning in activity recognition systems. In *4th European Conference on Smart Sensing and Context*, pages 121–134, 2009.

[93] A. Calatroni, D. Roggen, and G. Tröster. Automatic transfer of activity recognition capabilities between body-worn motion sensors: Training newcomers to recognize locomotion. In *Proc. 8th Int Conf on Networked Sensing Systems*, 2011.

[94] O. Banos, M. Damas, H. Pomares, F. Rojas, B. Delgado-Marquez, and O. Valenzuela. Human activity recognition based on a sensor weighting hierarchical classifier. *Soft Computing*, 17:333–343, 2013.

[95] S. Theodoridis and K. Koutroumbas. *Pattern Recognition*. Academic Press, 4th edition, 2008.

[96] J. A. Ward, P. Lukowicz, and H. W. Gellersen. Performance metrics for activity recognition. *ACM Trans. Intell. Syst. Technol.*, 2(1):6:1–6:23, 2011.

[97] X. Sun, H. Kashima, and N. Ueda. Large-scale personalized human activity recognition using online multitask learning. *IEEE Transactions on Knowledge and Data Engineering*, 25(11):2551–2563, 2013.

[98] J. Wu, A. Osuntogun, T. Choudhury, M. Philipose, and J.M. Rehg. A scalable approach to activity recognition based on object use. In *IEEE 11th International Conference on Computer Vision*, pages 1–8, 2007.

[99] C.V.C. Bouten, K.T.M. Koekkoek, M. Verduin, R. Kodde, and J.D. Janssen. A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. *IEEE Transactions on Biomedical Engineering*, 44(3):136–147, 1997.

[100] M. J. Mathie, A. C. F. Coster, N. H. Lovell, and B. G. Celler. Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement. *Physiological Measurement*, 25(2):1–20, April 2004.

[101] O. Banos, M. Damas, H. Pomares, A. Prieto, and I. Rojas. Daily living activity recognition based on statistical feature quality group selection. *Expert Systems with Applications*, 39(9):8013–8021, 2012.

[102] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification (2nd Edition)*. Wiley-Interscience, 2000.

[103] T. M. Cover and P. E. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27, 1967.

[104] N. Cristianini and J. Shawe-Taylor. *An introduction to support Vector Machines: and other kernel-based learning methods*. Cambridge University Press, New York, NY, USA, 2000.

[105] D. Roggen, A. Calatroni, M. Rossi, T. Holleczek, K. Förster, G. Tröster, P. Lukowicz, D. Bannach, G. Pirkl, A. Ferscha, J. Doppler, C. Holzmann, M. Kurz, G. Holl, R. Chavarriaga, M. Creatura, and J. del R. Millán. Collecting complex activity data sets in highly rich networked sensor environments. In *7th Int. Conf. on Networked Sensing Systems*, pages 233–240, 2010.

[106] T. Stiefmeier, D. Roggen, G. Ogris, P. Lukowicz, and G. Tröster. Wearable activity tracking in car manufacturing. *IEEE Pervasive Computing*, 7(2):42–50, April 2008.

[107] F. De la Torre, J. Hodgins, J. Montano, S. Valcarcel, R. Forcada, and J. Macey. Guide to the carnegie mellon university multimodal activity (cmu-mmac) dataset. Technical Report CMU-RI-TR-08-22, Robotics Institute, Carnegie Mellon University, July 2009.

[108] K. Frank, M. J. Vera, P. Robertson, and T. Pfeifer. Bayesian recognition of motion related activities with inertial sensors. In *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*, pages 445–446, 2010.

[109] Y. Xue and Lianwen J. A naturalistic 3d acceleration-based activity dataset and benchmark evaluations. In *IEEE International Conference on Systems Man and Cybernetics*, pages 4081–4085, Oct 2010.

[110] A. Reiss and D. Stricker. Creating and benchmarking a new dataset for physical activity monitoring. In *Proceedings of the 5th International Conference on PErvasive Technologies Related to Assistive Environments*, pages 40:1–40:8, 2012.

[111] Xsens Technologies B.V. *XM-B Technical Documentation*, May 2009. http://www.xsens.com.

[112] D. Bannach, O. Amft, and P. Lukowicz. Rapid prototyping of activity recognition applications. *IEEE Pervasive Computing*, 7(2):22–31, 2008.

[113] J. Sjöberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P.-Y. Glorennec, H. Hjalmarsson, and A. Juditsky. Nonlinear black-box modeling in system identification: a unified overview. *Automatica*, 31(12):1691–1724, 1995.

[114] L. Ljung. *System identification: theory for the user.* Prentice Hall, 1987.

[115] G. Box, G. M. Jenkins, and G. C Reinsel. *Time Series Analysis: Forecasting and Control, 3rd ed.* Prentice-Hall, 1994.

[116] O. Nelles. *Nonlinear System Identification.* Springer, 2000.

[117] X. Hong, R. J. Mitchell, S. Chen, C. J. Harris, K. Li, and G. W. Irwin. Model selection approaches for non-linear system identification: a review. *Intern. J. Syst. Sci.*, 39(10):925–946, October 2008.

[118] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods, 2nd ed.* Springer, 2009.

[119] B. L. Ho and R. E. Kalman. Efficient construction of linear state variable models from input/output functions. *Regelungstechnik*, 14:545–548, 1966.

[120] A. Yazdizadeh and K. Khorasani. Adaptive time delay neural network structures for nonlinear system identification. *Neurocomputing*, 47(1–4):207–240, 2002.

[121] M. Espinoza, J. A. K. Suykens, R. Belmans, and B. De Moor. Using kernel-based modeling for nonlinear system identification. *IEEE Control Systems Magazine*, pages 43–57, 2007.

[122] A. V. Vecchia. Periodic autoregressive-moving average (PARMA) modeling with applications to water resources. *Water resources bulletin*, 21(5):721–730, 1985.

[123] S. D. Fassois. MIMO LMS-ARMAX identification of vibrating structures - part i: The method. *Mechanical Systems and Signal Processing*, 15(4):737–758, 2001.

[124] H. R. Pota. Mimo systems-transfer function to state-space. *IEEE Transactions on Education*, 39(1):97–99, February 1996.

[125] D. J. Berndt and J. Clifford. Using dynamic time warping to find patterns in time series. In *AAAI - KDD Workshop*, pages 359–370, 1994.

[126] L. Ye and E. Keogh. Time series shapelets: a novel technique that allows accurate, interpretable and fast classification. *Data Mining and Knowledge Discovery*, 22(1-2):149–182, 2011.

[127] H. Harms, O. Amft, R. Winkler, J. Schumm, M. Kusserow, and G. Troester. Ethos: Miniature orientation sensor for wearable human motion analysis. In *Proc. of the IEEE Sensors Conference*, pages 1037–1042. IEEE, 2010.

[128] J. Sung, C. Ponce, B. Selman, and A. Saxena. Human activity detection from rgbd images. In *AAAI Workshop: Plan, Activity, and Intent Recognition*, 2011.

[129] E. Stone and M. Skubic. Evaluation of an inexpensive depth camera for passive in-home fall risk assessment. In *Proc Pervasive Health Conference*, pages 71–77, 2011.

[130] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from a single depth image. In *Proc of IEEE Conf on Computer Vision and Pattern Recognition*, pages 1297–1304, 2011.

[131] Prime sensor NITE 1.3 algorithms notes, version 1.0, 2010. http://www.primesense.com.

[132] http://code.google.com/p/qtkinectwrapper/.

[133] Xsens Technologies B.V. *XM-B Technical Documentation*, May 2009. http://www.xsens.com.

[134] S. Kallio, J. Kela, P. Korpipää, and J. Mäntyjärvi. User independent gesture interaction for small handheld devices. *International Journal of Pattern Recognition and Artificial Intelligence*, 20(4):505–524, 2006.

[135] T. Hongyong and Y. Youling. Finger tracking and gesture recognition with kinect. In *IEEE 12th International Conference on Computer and Information Technology*, pages 214–218, 2012.

[136] K. Förster, A. Biasiucci, R. Chavarriaga, J. del R. Millan, D. Roggen, and G. Tröster. On the use of brain decoded signals for online user adaptive gesture recognition systems. In *Proc. 8th Int. Conf. on Pervasive Computing*, pages 427–444, 2010.

[137] M. Berchtold, M. Budde, D. Gordon, H. Schmidtke, and M. Beigl. Actiserv: Activity recognition service for mobile phones. In *Proc. 14th Int. Symp. on Wearable Computers*, 2010.

[138] Y. Oussar and G. Dreyfus. How to be a gray box: dynamic semi-physical modeling. *Neural Networks*, 14(9):1161–1172, 2001.

[139] D. Roggen, A. Calatroni, K. Förster, G. Tröster, P. Lukowicz, D. Bannach, A. Ferscha, M. Kurz, G. Hölzl, H. Sagha, H. Bayati, J. del R. Millan, and R. Chavarriaga. Activity recognition in opportunistic sensor environments. *Procedia Computer Science*, 7:173–174, 2011.

[140] N. Kyriazis I. Oikonomidis and A.A. Argyros. Full dof tracking of a hand interacting with an object by modeling occlusions and physical constraints. In *Proceedings of the 13th IEEE International Conference on Computer Vision*, 2011.

[141] C. Rougier, E. Auvinet, J. Rousseau, M. Mignotte, and J. Meunier. Fall detection from depth map video sequences. In *Toward Useful Services for Elderly and People with Disabilities*, volume 6719, pages 121–128. Springer, 2011.

[142] J. C. Bongard and H. Lipson. Nonlinear system identification using coevolution of models and tests. *IEEE Transactions on Evolutionary Computation*, 9(4):361–384, 2005.

[143] N. Andriluka, S. Roth, and B. Schiele. Monocular 3d pose estimation and tracking by detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 623–630, 2010.

[144] O. Banos, M. Damas, H. Pomares, and I. Rojas. Handling displacement effects in on-body sensor-based activity recognition. In *5th International Work-conference on Ambient Assisted Living an Active Ageing*, volume 8277 of *Lecture Notes in Computer Science*, pages 80–87, 2013.

Curriculum Vitae

Personal Information

Oresti Baños
Born Juni 29, 1986, Hannover, Germany
Citizen of Spain

Education

2010 - 2014	PhD studies at CASIP group, Department of Computer Architecture and Computer Technology, University of Granada, Spain
2009 - 2011	MsC Electronic Engineering at Facultad de Ciencias, University of Granada, Spain
2009 - 2010	Master's Degree in Computer and Network Engineering (Postgraduate Course), University of Granada, Spain
2004 - 2009	M.Sc. Telecommunication Engineering at E. T. S. Ingenieras Informática y de Telecomunicación, University of Granada, Spain
2002 - 2004	High school at IES Montes Orientales, Iznalloz, Granada, Spain
1992 - 2002	Primary and Secondary School at Hermanos Coronel Velázquez, Píñar, Granada, Spain

Research/Work Experience

2010 - Today	Research Assistant at CASIP group, Department of Computer Architecture and Computer Technology, University of Granada, Spain
2010 - Today	National Research Grant Holder (FPU Fellowship, AP-2009-2244), Granada, Spain
2012 - 2012	Research Assistant at Activity and Context recognition Technologies Lab. (ACTLab), Department of Signal Processing Systems, Technical University of Eindhoven (TUe), Eindhoven, Netherlands
2011 - 2011	Research Assistant at Wearable Computing Lab., Department of Information Technology and Electrical Engineering, ETH Zürich, Switzerland
2011 - 2011	Visiting Researcher at Computer Laboratory of Ambient and Wearable Systems (CLAWS), Department of Electrical and Computer Engineering, University of Alabama, Tuscaloosa, USA
2010 - 2011	University of Granada Research Fellowship, Granada, Spain
2010 - 2010	Goverment of Andalusia Research Grant Holder (Project of Excellence, P09-TIC-175476), Granada, Spain

Publications List

International Journals (SCI-indexed)

Banos, O., Damas, M., Guillen, A., Herrera, L.J., Pomares, H., Rojas, I. Multi-sensor fusion based on asymmetric decision weighting for robust activity recognition. **Neural Processing Letters**, *Springer* (2014) [Under review]

Banos, O., Galvez, J. M., Damas, M., Pomares, H., Rojas, I. Window size impact in activity recognition. **Sensors**, *MDPI.* (2013) [Under review]

Banos, O., Damas, M., Gloesekoetter, P., Pomares, H., Rojas, I. PhysioDroid: combining wearable health sensors and smartphones for a pervasive, continuous and personal monitoring. **Artificial Intelligence in Medicine**, *Elsevier* (2013) [Under review]

Banos, O., Damas, M., Pomares, H., Rojas, F., Delgado-Marquez, B., Valenzuela, O. Human activity recognition based on a sensor weighting hierarchical classifier. **Soft Computing**, *Springer*, vol. 17, pp. 333-343 (2013)

Banos, O., Damas, M., Pomares, H., Rojas, I. On the Use of Sensor Fusion to Reduce the Impact of Rotational and Additive Noise in Human Activity Recognition. **Sensors**, *MDPI*, vol. 12, no. 6, pp. 8039-8054 (2012)

Banos, O., Damas, M., Pomares, H., Prieto, A., Rojas, I.: Daily Living Activity Recognition based on Statistical Feature Quality Group Selection. **Expert Systems with Applications**, *Elsevier*, vol. 39, no. 9, pp. 8013-8021 (2012)

Book Chapters

Banos, O., Toth M. A., Damas, M., Pomares, H., Rojas, I. Amft, O.: Evaluation of inertial sensor displacement effects in activity recognition systems. Science and Supercomputing in Europe (Information & Communication Technologies), **HPC-Europe 2** (2013) ISBN: 978-84-338-5400-1

International Conferences

Banos, O., Damas, M., Pomares, H., Rojas, I.: Handling displacement effects in on-body sensor-based activity recognition. In: Proceedings of the 5th International Work-conference on Ambient Assisted Living an Active Ageing (**IWAAL 2013**), San Jose, Costa Rica, December 2-6, (2013) [*BEST PAPER AWARD*]

Herrera L. J., Rojas, I., Pomares, H., Guillen A., Valenzuela O.: Classification of MRI images for Alzheimer's disease detection. In: Proceedings of the ASE-IEEE International Conference on on Biomedical Computing (**BioMedCom 2013**), Washington D.C., USA, September 8-14, (2013)

Valenzuela, O., Rojas, F., Herrera, L. J., Ortuno, F., Banos, O., Ruiz, G., Tribak, H., Pomares, H., Rojas, I.: Intelligent systems to autonomously classify several arrhythmia using information from ECG. In: Proceedings of the ASE-IEEE International Conference on Biomedical Computing (**BioMedCom 2013**), Washington D.C., USA, September 8-14, (2013)

Banos, O., Damas, M., Pomares, H., Rojas, I.: Activity recognition based on a multi-sensor meta-classifier. In: Proceedings of the 2013 International Work Conference on Neural Networks (**IWANN 2013**), Tenerife, June 12-14, (2013)

Banos, O., Damas, M., Gloesekoetter, P., Hermes, A., Hendrik, H., Pomares, H., Rojas, I.: PhysioDroid: an app for physiological data monitoring. In: Proceedings of the International Work-Conference on Bioinformatics and Biomedical Engineering (**IWBBIO 2013**), Granada, Spain, March 18-20, (2013)

Banos, O., Toth, M. A., Damas, M., Pomares, H., Rojas, I., Amft, O.: A benchmark dataset to evaluate sensor displacement in activity recognition. In: Proceedings of the 14th International Conference on Ubiquitous Computing (**Ubicomp 2012**), Pittsburgh, USA, September 5-8, (2012)

Banos, O., Calatroni, A., Damas, M., Pomares, H., Rojas, I., Troester, G., Sagha, H., Millan, J. del R., Chavarriaga, R., Roggen, D.: Kinect=IMU? Learning MIMO Signal Mappings to Automatically Translate Activity Recognition Systems Across Sensor Modalities. In: Proceedings of the 16th annual International Symposium on Wearable Computers (**ISWC 2012**), Newcastle, United Kingdom, June 18-22 (2012)

Banos, O., Damas, M., Pomares, H., Rojas, I.: Human multisource activity recognition for AAL problems. In: Proceedings of the 5th International Symposium on Ubiquitous Computing and Ambient Intelligence (**UCAmI 2011**), Riviera Maya, Mxico, December 5-9, (2011)

Banos, O., Damas, M., Pomares, H., Rojas, I.: Application of a novel feature selector for human activity recognition based on inertial monitored data. In: Proceedings of the 2011 International Conference on Artificial Intelligence (**ICAI 2011**), Las Vegas, Nevada, July 18-21, (2011)

Banos, O., Damas, M., Pomares, H., Rojas, I.: Recognition of Human Physical Activity based on a novel Hierarchical Weighted Classification scheme. In: Proceedings of the 2011 International Joint Conference on Neural Networks (**IJCNN 2011**), IEEE, San Jos, California, July 31-August 5, (2011)

Banos, O., Damas, M., Pomares, H., Rojas, I.: Automatic Recognition of Daily Living Activities based on a Hierarchical Classifier. In: Proceedings of the 2011 International Work Conference on Neural Networks (**IWANN 2011**), Torremolinos, Mlaga, June 8-10, (2011)

Banos, O., Pomares, H., Rojas, I.: Ambient Living Activity Recognition based on Feature-set Ranking Using Intelligent Systems. In: Proceedings of the 2010 International Joint Conference on Neural Networks (**IJCNN 2010**), IEEE, Barcelona, July 18-23, (2010)

Banos, O., Pomares, H., Rojas, I.: Novel Method for Feature-set Ranking Applied to Physical Activity Recognition. 23rd International Conference on Industrial Engineering and Other Applications of Applied Intelligent Systems (**IEA/AIE 2010**), Cordoba, June 1-4, pp. 637-642 (2010)

Supervised M.Sc. Projects

Computer Science

Dec 2013 *Title:* mHealthDroid: A toolkit for rapid development of mHealth applications. *Author:* Alejandro Saéz. *Institution:* University of Granada, Spain

Dec 2013 *Title:* Design, implementation and validation of a mobile framework for agile development of mHealth applications. *Author:* Rafael García. *Institution:* University of Granada, Spain

Sep 2012 *Title:* Sensmotion: APP for daily living activities monitoring. *Author:* Raúl Nadal. *Institution:* University of Granada, Spain

Sep 2011 *Title:* Human Activity Recognition based on an application for the Android operating system. *Author:* Jens Ostrowski. *Institution:* University of Applied Sciences Mnster, Germany

Electronic Engineering

Sep 2013 *Title:* On the analysis of window-size importance in activity recognition. Increasing systems reliability through multi-segmented fusion. *Author:* Juan Manuel Gálvez. *Institution:* University of Granada, Spain

Oct 2012 *Title:* Android APP for biomedical monitoring. *Author:* Hendrik Mende. *Institution:* University of Applied Sciences Muenster, Germany

Oct 2012 *Title:* Development of an Android application for the monitoring, storage and upload of externally collected biomedical data. *Author:* Andreas Hermes. *Institution:* University of Applied Sciences Muenster, Germany

Telecommunications Engineering

Mar 2014 *Title:* Automatic detection of on-body sensors location for opportunistic and robust inference of behavior. *Author:* Marta Calles. *Institution:* University of Granada, Spain

Sep 2012 *Title:* Signal preprocessing effects on daily living activity recognition systems. Aplication to energy expenditure analysis. *Author:* Juan Manuel Gálvez. *Institution:* University of Granada, Spain

Project Contributions

European Projects

Wireless Health Monitoring (WHM)

Partners: Muenster University of Applied Sciences, Vector Fabrics, Centric TSolve, University Twente, Medische Spectrum Twente, TU Braunschweig.

Founded by: Euregio, European Regional Development Fund.

Project Duration: 2010-2013.

Reference ID: INTERREG IV, I-1-02=091.

HPC-Europa2

Partners: UNIFOB, CSC, EPCC, TCD, SARA, GENCI, HLRS, PSNC, CINECA, ICCS, BSC.

Founded by: European Commission.

Project Duration: 2010-2012.

Reference ID: EU-FP7 grant #228398.

Activity and Context Recognition with Opportunistic Sensors Configurations (OPPORTUNITY)

Partners: Swiss Federal Institute of Technology, University of Passau, Johannes Kepler Universitt Linz and cole Polytechnique Fdrale de Lausanne.

Founded by: European Commission.

Project Duration: 2009-2012.

Reference ID: EU-FP7-FET-Open grant #225938.

National Projects

Development of advanced intelligent systems in high performance platforms. Application to bioinformatics and biomedical problems (DSIPA-BIO)

Partners: University of Granada.

Founded by: Ministry of Science and Innovation.

Project Duration: 2010-2013. *Reference ID:* SAF2010-20558.

Digital and Personal Environment for Health and Welfare (AmIVital)

Partners: Siemens, Telefonica I+D, Telvent, Ericsson, Eptron, CPI,

Acerca Comunicaciones, Airzone e IS2, ITACA (Grupo TSB), CARTIF Foundation, Rioja Salud Foundation, Puerta de Hierro University Hospital, Carlos III Health Institute, University of Granada, University of Malaga, Technical University of Madrid and University of Zaragoza.
Founded by: Center for Industrial Technological Development (CDTI).

Project Duration: 2007-2010.

Reference ID: CENIT2007-1010.

Regional Projects

High performance computing in bioinformatics and biomedicine by using intelligent systems (HPC-BIO)

Partners: University of Granada.

Founded by: Regional Ministry of Science and Innovation.

Project Duration: 2010-2013.

Reference ID: P09-TIC-175476 .

Advanced computing for dynamic, adaptive and self-organizing control of complex systems. Application to industrial and biomedical problems (ADA-BIO)

Partners: University of Granada.

Founded by: Regional Ministry of Science and Innovation.

Project Duration: 2008-2011.

Reference ID: P07-TIC-02768.