SAPIENZA
UNIVERSITÀ DI ROMA

Techniques for Indoor Localization and Activity Recognition

Dipartimento di Ingegneria Informatica Automatica e Gestionale ANTONIO RUBERTI,
SAPIENZA – Università di Roma

Dottorato di Ricerca in Ingegneria Informatica – XXVI Ciclo

Candidate

Francesco Leotta
ID number 799650

Thesis Advisor

Ing. Massimo Mecella

A report submitted in partial fulfillment of the requirements
for the admission to the second year of the Ph.D course in
Computing Science and Engineering

October 2011

Francesco Leotta. *Techniques for Indoor Localization andActivity Recognition.*
Ph.D. report. Sapienza – University of Rome

VERSION: 2011, October the 24th
WEBSITE: http://www.dis.uniroma1.it/~leotta
EMAIL: leotta@dis.uniroma1.it

Chapter 1

Fact Sheet

1.1 Introduction

This preliminary chapter briefly exposes the results that I obtained up to now. Section 1.2 describes the amount of credits I collected during the year. Section 1.3 reports the activities achieved and the publications that I contributed to.

1.2 Exams

Here follow table 1.1, summing up the exams that I already passed, and table 1.2 which indicates the B type courses that are going to be passed soon.

Table 1.1. This table summarizes the exams that were passed up to now.

Type	Name	Professor	Cr.	Date
A			12	
	Artificial Intelligence II	D. Nardi	6	2011-09-13
	Machine Learning	L. Iocchi	6	2011-10-17
C			7.5	
	Artifact Verification	G. De Giacomo	0.5	2011-09-29
	Business Artifacts with Lifecycles: Motivations, Applications, Direction	R. Hull	0.5	2011-09-29
	DISC Workshops and Tutorials	Various	4	2011-09-23
	Learn to Behave	V. A. Ziparo	0.5	2011-06-06
	Rescue Robotics (Sensing and State Estimation)	A. Kleiner	0.5	2011-04-01
	Seminar on Nonlinear system identification and sensor fusion	T. Schoen	0.5	2011-05-19
	Seminar on Windows Azure	A. Musone	0.5	2011-04-14
	The “Roman Model” for Automatic Synthesis in Practice: the SM4All Experience	C. Di Ciccio	0.5	2011-07-26

Table 1.2. This table summarizes the exams that have been already frequented but not yet passed.

Type	Name	Professor	Credits
B			
	Supervisory Control of Discrete-Event Systems	W. M. Wonham	2.5
	Least Squares Estimation and SLAM	G. Grisetti	2.5

1.3 Main activities and publications

1.3.1 Research project activities

The TOBI project

I have been actively involved in Work Packages 1, 5 and 8 of the TOBI[1] European Research Project (http://www.tobi-project.org/) whose aim is to develop practical technology for brain-computer interaction (BCI) that will improve the quality of life of disabled people and the effectiveness of rehabilitation.

A Brain Computer Interface (BCI) allows to use the electroencephalographic (EEG) signal of a subject as control signal for computer applications extracting some features. My work in the TOBI project has been mainly focused on the following aspects:

- **Standardization**. Though BCI is an active research field since '90s, there not exists any standardized way of making modular BCI systems due to the lack of standard interfaces for the transmission of raw biosignals, signal features, classification results, markers and events. My work in TOBI WP8 has been directed at the definition of such a kind of interfaces. Additionally, as a proof of concept, I have used the defined interface to design and develop a portable BCI system.

- **Hybrid BCI**. Traditional BCIs extract control features using a sole EEG phenomenon. TOBI WP5 aims at defining an Hybrid BCI able to exploit different EEG phenomenon at the same time and to optionally combine them with other kind of bio-signals (such as electromyogram - EMG and electrooculogram - EOG). My work on this work package has been focused on the implementation of the interfaces defined in WP8.

- **Combination of BCI with established Assistive Technology (AT) platforms**. My work on this work package has been focused on the design of a BCI P300 stimulation to be overlaid over existing AT softwares.

Publications I contributed to the writing of 6 papers, related to the TOBI project, which are:

- F. Aloise, F. Schettini, P. Aricò, F. Leotta, S. Salinari, D. Mattia, F. Babiloni and F. Cincotti. P300-based BCI for Environmental Control: an Asynchronous Approach. *Journal of Neural Engineering*, 2011.

- A. Riccio, F. Leotta, L. Bianchi, F. Aloise, C. Zickler, EJ. Hoogerwerf, A. Kubler, D. Mattia and F. Cincotti. Workload Measurement in a Communication Application Operated through a P300-based BCI. *Journal of Neural Engineering*, 2011.

- F. Aloise, F. Schettini, P. Aricò, F. Leotta, S. Salinari, F. Babiloni, D. Mattia and F. Cincotti. Towards Domotic Appliances Control through a Self-Paced P300-based BCI. *International Conference on Bio-Inspired Systems and Signal Processing (BIOSIGNAL)*, 2011.

- F. Schettini, F. Aloise, P. Aricò, F. Leotta, S. Salinari, F. Babiloni, D. Mattia and F. Cincotti. Improving Asynchronous Control for P300-based BCI: towards a completely autoadaptative system. *In Proceedings of 2nd Workshop of the TOBI Project*, 2010.

- GR. Muller-Putz, C. Breitwieser, M. Tangermann, M. Schreuder, M. Tavella, R. Leeb, F. Cincotti, F. Leotta and C. Neuper. TOBI Hybrid BCI: Principle of a New Assistive Method. *In Proceedings of 2nd Workshop of the TOBI Project*, 2010.

- SC. Kleih, T. Kaufmann, C. Zickler, S. Halder, F. Leotta, F. Cincotti, F. Aloise, A. Riccio, C. Herbert, D. Mattia and A. Kubler. Out of the frying pan into the fire - the P300 based BCI faces real world challenges. Book chapter. *Progress in Brain Research*, Vol. 194, 2011.

[1]Tools for Brain Computer Interfaces

Chapter 2

Research

2.1 Introduction

Last years witnessed a growing interest in providing intelligent services for both indoor and outdoor environments. The list of fields of application that can take advantage of this kind of services includes:

- Ambient Assisted Living (AAL) which aims at providing supervision or assistance with activities of daily living. AAL systems monitor resident activities in order to provide help or automation to ensure their health, safety, and well-being.
- Building Automation. Awareness of mobile agents activities and habits could aim at devising intelligent services including those intended for power saving strategies.
- Security monitoring for homes, warehouses, banks, stores, airports, hospitals, etc.. Access control for example represents a typical agent monitoring application allowing for logging allowed access to some kind of resources and detecting unexpected or disallowed access. Last years have shown an emerging interest in the detection of suspect or dangerous behavior by mean of activity analysis [HBC+05].
- Traffic monitoring [BI09].
- Position of emergency squads into buildings.

Ambient awareness represents a common necessity for these services whose behavior depends, in particular, on environmental dynamics. In the vast majority of cases mobile agents, and people in particular, are central in environmental changes, so their positions, poses and activities are crucial at devising well-designed intelligent services.

Systems able to gather information about users moving in a certain environment can be considered as low level blocks for the above mentioned high level intelligent services. In particular these systems can provide basically two kinds of low level (in a hierarchical sense) services:

- Location and Pose Sensing Services (in the following simply Localization Services). Services belonging to this category provide information about the positions, and sometimes the poses, of mobile agents detected into the environment. Sometimes they also try to provide information about agents identities. This association could be straightforward in the case for marker-based systems while it could be more tricky for marker-less systems (see 2.2.1 for details).
- Activity Recognition Services. They try to classify mobile agents actions and activities.

The choice of a certain implementation of a low level service system strongly depends on the final application. For example, a home automation system in some case does not need an highly precise location sensing system, gathering only rough information about room occupancy; more complex AAL systems may require an higher degree of precision providing services that are strongly dependent on users exact positions and poses. In addition, the cost (in terms of money and computational or energy requirements) of a system implementing a low level service is likely dependent from its performances. So, the choice of a specific implemented system can be reduced to a trade-off between the needed performance and the cost for the customer. Recent advances in off-the-shelf technology and equipment (Microsoft Kinect [SFC+11] represents a good example) will likely enlarge the list of available solutions.

2.2 Related Works: Location and Pose Sensing services

A localization system is in charge of providing services which aim at providing information about dynamical positions (e.g. trajectory), dynamical poses and identity of mobile agents detected into the monitored environment. In particular, an implemented system could provide only a subset of the above mentioned features with different levels of accuracy for each of them.

Usual performance measures for trajectory estimation are *accuracy* and *precision*. Accuracy (or location error) is the most important requirement of positioning systems. Usually, mean distance error is adopted as the performance metric, which is the average Euclidean distance between the estimated location and the true location. Accuracy can be considered to be a potential bias, or systematic effect/offset of a positioning system. Accuracy only considers the value of mean distance errors. However, location precision considers how consistently the system works, i.e., it is a measure of the robustness of the positioning technique as it reveals the variation in its performance over many trials.

A localization service can provide two kinds of information: *physical* and *symbolic*. In contrast to physical a location, symbolic location encompasses abstract ideas of where something is: in the kitchen, next to a door, to the fridge and so on. Using symbolic locations instead can be considered as a way to add semantic to physical position and this feature can be useful to activity recognition services as well as to coordination services [JLY10] [CIN+09].

Also pose estimation allows for different degree of accuracy. The accuracy in particular may range from simple bounding volume to accurate estimations using geometric body models [PI08].

A localization system may be able to locate objects worldwide, within a metropolitan area, throughout a campus, in a particular building, or within a single room. Further, the number of objects the system can locate with a certain amount of infrastructure or over a given time may be limited. For example, GPS can serve an unlimited number of receivers worldwide using 24 satellites plus three redundant backups. On the other hand, some electronic tag readers cannot read any tag if more than one is within range. To assess the scale of a location-sensing system, we consider its coverage area per unit of infrastructure and the number of objects the system can locate per unit of infrastructure per time interval. This chapter will introduce mainly systems for **indoor and limited outdoor** environments while systems for geolocation will be briefly introduced in section 2.2.2.

2.2.1 Marker-based and Marker-less approaches

A first rough classification divides localization systems into those which require the mobile agents to wear special markers or tags (*marker-based systems*) and those which suppose mobile agents passively participating to the ongoing localization task not requiring any body area tag (*marker-less systems*). Because of this definition the terms *active* and *passive* respectively are used to denote the two kinds of location systems.

While passive location systems are easily recognizable as less invasive from the point of view of the monitored agents (especially if they are humans), their complexity is usually higher with respect to active systems. In particular a passive system has to solve a set of problem that are immediately addressed in active systems:

- **Background modeling**. Usually passive systems start with separating the so called *foreground* composed by mobile agents to monitor from the rest of the environment *background*. This task is not trivial due to the fact that background usually change (though with a low rate) and can sometimes contain periodically moving elements (tree leaves, curtains, etc.). Depending on the technology used, as we will see into further details exploring already proposed solutions, additional issues can make this task even harder. Additionally background extraction is followed by a **segmentation** phase, which consists of identifying mobile agents into the foreground, that may become difficult in **crowded** environments. A review of techniques used for background modeling with different sensing technology can be found in [CFBM10].

- **Tracking**. In active systems, usually, each tag has a unique ID. So, provided that the system is able to calculate frame by frame the position of a tag, it is possible to follow the trajectory of a mobile agent at each time step simply grouping measurements relative to a certain tag ID. In passive systems the association between agents detected at time t and those already tracked at time $t-1$ is more complicated implying the extraction and comparison of a set of features dependent from the sensing technology.

- **Recognition**. Sometimes a location system aims at recognizing the identity of the detected mobile agents. Once again this task can be simple in active systems (provided that the association between tag ID and mobile agents identity is known) while could require sophisticated techniques in the case of passive systems especially because users are not always truthful (for example in surveillance applications).

On the other side active systems present the following limitations with respect to passive systems:

- Because of the fact that each mobile agent is usually provided with a sole tag, in the vast majority of the cases, marker-based systems provide only the position of a mobile agent tag in the space and cannot provide information about the pose. In order to use marker-based systems for pose estimation, the mobile agents has to wear a huge number of tag; this strategy allows to obtain a very high precision and is used in gaming industry for motion capture (see MotionStart DC magnetic tracker system for example) but it is, in general, very expensive and so it has a limited applicability.

- Active systems can be used only if case the system has to deal only with well known agents as in the case of AAL and building automation applications. In general this is not true, so some kind of applications simply cannot relay on active systems. This is the case, for example, of monitoring applications for traffic analysis or security.

2.2.2 Wireless Approaches for Localization Services

Wireless localization systems are usually **marker-based**. They are mainly based on the transmission of signals between agents tags and readers or between tags. Triangulation and proximity are the two principal techniques for location sensing with a wireless medium.

The triangulation location sensing technique uses the geometric properties of triangles to compute object locations. Triangulation is divisible into the sub-categories of *lateration*, using distance measurements, and *angulation*, using primarily angle or bearing measurements.

We define the term lateration to mean for distance measurements what angulation means for angles. Lateration computes the position of an object by measuring its distance from multiple reference positions. Calculating an object's position in two dimensions requires distance measurements from 3 non-collinear points whilst in 3 dimensions, distance measurements from 4 non-coplanar points are required. Domain-specific knowledge may reduce the number of required distance measurements [HHS$^+$02].

There are three general approaches to measuring the distances required by the lateration technique.

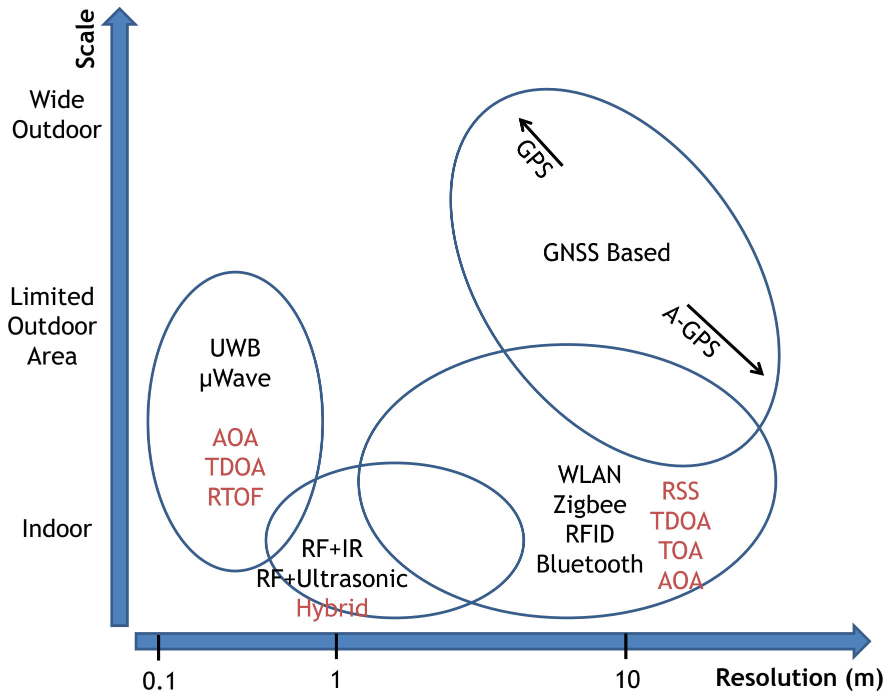

Figure 2.1. Representation of main marker-based wireless approaches for localization inf the Resolution-Scale plane.

- **Time-of-Flight**. Measuring distance from an object to some point P using time-of-flight means measuring the time it takes to travel between the object and point P at a known velocity.
- **Attenuation**. The intensity of an emitted signal decreases as the distance from the emission source increases. The decrease relative to the original intensity is the attenuation.
- **Phase of Arrival**.

Angulation is similar to lateration except, instead of distances, angles are used for determining the position of an object. In general, two dimensional angulation requires two angle measurements and one length measurement such as the distance between the reference points. In three dimensions, one length measurement, one azimuth measurement, and two angle measurements are needed to specify a precise position.

A proximity location-sensing technique entails determining when an object is “near” a known location. The object’s presence is sensed using a physical phenomenon with limited range. There are three general approaches to sensing proximity namely **detecting physical contact** (by means of pressure sensors, touch sensors, and capacitive field detectors), **monitoring wireless cellular access points** (monitoring when a mobile device is in range of one or more access points in a wireless cellular network) and **observing automatic ID systems** (such as credit card point-of-sale terminals, computer login histories, telephone records, electronic card lock logs, and identification tags such as electronic highway E-Toll systems). In particular, monitoring wireless cellular access point is frequently used in wireless based systems [JLY09].

Figure 2.1 shows how the different techniques are used in combination to the more extensively used wireless technologies namely:

- **RFID** is a means of storing and retrieving data through electromagnetic transmission to an RF compatible integrated circuit and is now being seen as a means of enhancing data handling processes. An RFID system has several basic components, including a number of

RFID readers, RFID tags, and the communication between them. The RFID reader is able to read the data emitted from RFID tags. RFID readers and tags use defined RF and protocol to transmit and receive data. RFID tags are categorized as either passive or active. Passive RFIDs are not provided with a battery and usually modulate the signal produced by the RFID reader, so they have a very limited range. Active RFIDs have their own on-board power source, so they have enough energy to produce long range signals and to empower more sophisticated circuits (including on-board memories). See dedicated subsection for details.

- **UWB** is based on sending ultrashort pulses (typically < 1 ns), with a low duty cycle (typically $1 : 1000$). On the spectral domain, the system, thus, uses an UWB (even >500 MHz wide). UWB location has the following advantages. Unlike conventional RFID systems, which operate on single bands of the radio spectrum, UWB transmits a signal over multiple bands of frequencies simultaneously, from 3.1 to 10.6 GHz. UWB signals are also transmitted for a much shorter duration than those used in conventional RFID. UWB tags consume less power than conventional RF tags and can operate across a broad area of the radio spectrum. UWB can be used in close proximity to other RF signals without causing or suffering from interference because of the differences in signal types and radio spectrum used. See dedicated subsection for details.

- The midrange wireless local area network (**WLAN**) standard, operating in the 2.4-GHz Industrial, Scientific and Medical (ISM) band, has become very popular in public hotspots and enterprise locations during the last few years. With a typical gross bit rate of 11, 54, or 108 Mbps and a range of 50-100 m, IEEE 802.11 is currently the dominant local wireless networking standard. It is, therefore, appealing to use an existing WLAN infrastructure for indoor location as well, by adding a location server. The accuracy of typical WLAN positioning systems using RSS is approximatly 3 to 30 m, with an update rate in the range of few seconds. See dedicated subsection for details.

- **Bluetooth** operates in the 2.4-GHz ISM band. Compared to WLAN, the gross bit rate is lower (1 Mbps), and the range is shorter (typically 10-15 m). On the other hand, Bluetooth is a “lighter” standard, highly ubiquitous (embedded in most phones, personal digital assistants (PDAs), etc.) and supports several other networking services in addition to IP. Bluetooth tags are small size transceivers. As any other Bluetooth device, each tag has a unique ID. This ID can be used for locating the Bluetooth tag. The Topaz[1] is an example of Bluetooth based location system providing an accuracy of 2m for 95% of the time.

- Infrared (**IR**) positioning systems have been widely used as positioning systems (and nowadays represents the most widely deployed class of location systems [WHFaG92]), because IR technology is available on board of various wired and wireless devices, such as TV, printer, mobile phones, PDAs, etc. An IR-based positioning system, which offers absolute position estimations, needs line-of-sight communication between transmitters and receivers without interference from strong light sources. Thus the coverage range per infrastructure device is limited within a room. Nowadays commercial solutions like Radianse[2] and Versus[3] exploit IR for fine grain location in combination with RF techniques for coarse-grain location.

- A **Ultrasound** time-of-flight (TOA) lateration technique is employed by Active Bat [HHS+02] location system. The system can locate tags to within 9cm of their true position for 95 percent of the measurements but on the other side it relies on an expensive and sensitive (to precise placement) grid of ceiling sensors. The HP Labs Smart-LOCUS uses ultrasounds in combination with RF.

[1] http://www.tadlys.com
[2] http://www.radianse.com
[3] http://www.versustech.com

- Proprietary **UHF** systems exploit the fact that at such frequencies ranges wall attenuation is moderate. It is frequent in particular that this kind of systems works at the same frequency band of WLAN (2.4 GHz) (WhereNet system for example is very similar as approach to WLAN-based RADAR (see section 2.2.2)) using proprietary protocols allowing for low power consumption.

RFID based systems

SpotON [HWB00] is a well-known location system using RFID technology. SpotON represents an example of *Ad Hoc Location Sensing* system where objects are located without relying on an infrastructure. In a purely ad hoc location-sensing system, all of the entities become mobile objects with the same sensors and capabilities. To estimate their locations, objects cooperate with other nearby objects by sharing sensor data to factor out overall measurement error. In this way, a cluster of ad hoc objects converges to an accurate estimate of all nearby objects' positions. Objects in the cluster are located relative to one another or absolutely if some objects in the cluster occupy known locations. The SpotON system implements ad hoc lateration with low-cost active tags. SpotON tags use radio signal attenuation to estimate intertag distance. They exploit the density of tags and correlation of multiple measurements to improve both accuracy and precision. Sensing object locations with no fixed infrastructure represents a highly scalable and low-cost approach. Infrastructural systems could incorporate ad hoc concepts to increase accuracy or reduce cost.

LANDMARC [NLLP04] represents an example of RFID *Infrastructure* based location system. LANDMARC implement a sort of three level infrastructure with:

- a set of **RFID readers**. Each of them providing the signal power of each of the RFID tags detected, with signal power expressed using numbers from 1 to 8.
- in order to reduce the number of RFID readers needed in order to obtain a certain accuracy, a set of **reference RFID tags** with known position is placed into the monitored environment.
- a set of **tracked RFID tags** that are located using the power level calculated by the RFID readers in the range compared to the power level of the reference tags.

Active tags used in LANDMARC can guarantee a very long battery life of 3-5 years considering an update interval of 7.5 secs. The system provide an accuracy <2m which is <1m for 50% of the time.

Both SpotON and LANDMARC makes use of active RFID tags. Passive RFID tags have a very limited reading range (5-8 cm) and so their use in location system is limited for measurement involving close contact interaction [TGLP08].

UWB based systems

A 3-tier UWB-based indoor localization system is described in [LDG09]. The system exploits TOA measurements of transmitting only tags. The signal emitted by each tag is received by a set of **hubs** which has not to be synchronized (the clock difference can be eliminated using multiple hub timestamps) that act as relay station for base stations which performs the computations about tags positions. The paper shows a direct relation between the accuracy of measuring hub locations and the accuracy of the measurements provided by the system; for example in order to obtain an accuracy of 10cm a 1 cm hub location accuracy must be guaranteed (the hub location can be setup manually or calculated at runtime using base stations).

Commercial systems based on UWB are also available. An example is Ubisense system[4].

[4] http://www.ubisense.net

WLAN based systems

WLAN approach as the main advantage of exploiting the same infrastructure that provides the building's general purpose wireless networking for location service. As a disvantage, the object it is tracking must support a wireless LAN, which may be impractical on small or power-constrained devices, this problem affects the vast majority of WLAN marker-based systems.

In 2000 Microsoft Research developed RADAR [BP00], a building-wide tracking system based on the IEEE 802.11 WLAN standard. Microsoft has developed two RADAR implementations: one using scene analysis (marker less) and the other using lateration (marker-based). The RADAR lateration approach is based on k-nearest neighbors (k-nn) and provide an accuracy of 4.3 meters with 50% of probability. The main problem with this system is that generalizing RADAR to multifloored buildings or three dimensions presents a nontrivial problem.

EKAHAU commercial system[5] follows instead of k-nn uses the probabilistic approach described in [RMT+02]. The approach is based (as in RADAR) on signal attenuation but the problem is modeled as a kernel based machine learning problem.

Wireless Marker-less systems

Wireless Marker-less approaches are in general less frequent in literature with respect to marker-based approaches.

An example of this kind of system is the already mentioned scene-analysis profile of RADAR system based on the standard WLAN infrastructure. The system provide an accuracy of 3 meters with a 50% probability (so better then the RADAR lateration profile) but as a drawback no mechanism is suggested for dealing with background dynamics.

Recently an approach based on a generic WSN infrastrucure have been proposed in [VRB+10]. Here the authors divide the floor into cells and than the association of an agent is transformed into an SVM problem of classification. The system shows an accuracy within 1m tested with two agents moving at the same time into the environment.

Global Localization Systems

Up to this moment, we focused or attention of positioning systems working in indoor or limited outdoor environment. In this section we will briefly discuss about **global** localization systems. Geolocation is closely related to positioning but can be distinguished from it by a greater emphasis on determining a meaningful location (e.g. a street address) rather than just a set of geographic coordinates.

Global Navigation Satellite Systems (GNSS) represents a well known class of global navigation systems. As of 2010, the United States NAVSTAR **Global Positioning System** (GPS) was the only fully operational GNSS. The Russian GLONASS achieved full global coverage in October 2011 after the successful launch of the latest GLONASS satellite. The People's Republic of China is in the process of expanding its regional Beidou navigation system into the global Compass navigation system by 2020. The European Union's **Galileo** positioning system is a GNSS in initial deployment phase, scheduled to be fully operational by 2020 at the earliest. The Indian Regional Navigational Satellite System (IRNSS) is an autonomous regional satellite navigation system being developed by Indian Space Research Organisation.

While each of the previous systems differs from some details, GPS can be used as example to describe the entire class. The nominal GPS Operational Constellation consists of 24 satellites that orbit the earth in 12 hours. There are often more than 24 operational satellites as new ones are launched to replace older satellites. The satellite orbits repeat almost the same ground track (as the earth turns beneath them) once each day. The orbit altitude is such that the satellites repeat the same track and configuration over any point approximately each 24 hours (4 minutes earlier

[5] http://www.ekahau.com

each day). There are six orbital planes (with nominally four satellites in each), equally spaced (60 degrees apart), and inclined at about fifty-five degrees with respect to the equatorial plane. This constellation provides the user with between five and eight satellites visible from any point on the earth.

Privacy is an important issue in global positioning field. In the vast majority of the cases the computation about position is performed directly over the agent (the GPS receiver) using TOA and this imply a tight synchronization between the elements of the infrastructure. However, in a system like GPS, the receiver is not synchronized with the satellite transmitters and thus cannot precisely measure the time it took the signal to reach the ground from space. Therefore, GPS satellites are precisely synchronized with each other and transmit their local time in the signal allowing receivers to compute the difference in time-of-flight. GPS receivers can compute their 3-dimensional position (latitude, longitude, and elevation) using 4 satellites. The satellites are always above the receivers so only 3 satellites would normally be required to provide distance measurements in order to estimate a 3D position. However in GPS a fourth satellite measurement is required to allow us to solve for the forth unknown, the error between the receiver clock and the synchronized satellite clocks.

Poor coverage of satellite signal for indoor environments makes GNSS systems unsuitable for indoor location estimation. Assisted GPS (A-GPS) systems (like SnapTrack[6]) where a reference GPS receiver is used to relay GPS signal reaching an accuracy of 5-50m. Alternatively GPS can be used for indoor location sensing coupling it with an **Inertial Navigation System** (INS).

An INS includes at least a computer and a platform or module containing accelerometers, gyroscopes, or other motion-sensing devices. The INS is initially provided with its position and velocity from another source (a human operator, a GPS satellite receiver, etc.), and thereafter computes its own updated position and velocity by integrating information received from the motion sensors. The advantage of an INS is that it requires no external references in order to determine its position, orientation, or velocity once it has been initialized. ENSCO[7] provide INS systems for disaster management and war operations.

Wi-Fi-based positioning systems (WPS) emerged as an idea that can solve the positioning in certain situations (also indoor), taking advantage of the rapid growth of wireless access points in urban areas. Skyhook Wireless is one provider of this type of service, maintaining a public database that can be accessed through an API, and get the position based on the access points are accessible from a terminal. Other providers include the Fraunhofer Institute or Google.

The localization technique used for positioning with wireless access points is based on measuring the intensity of the received signal (received signal strength RSS). The accuracy depends on the number of positions that have been entered into the database. The possible signal fluctuations that may occur can increase errors and inaccuracies in the path of the user.

A number of systems have used global system of mobile/code division multiple access (GSM/CDMA) mobile cellular network to estimate the location of outdoor mobile clients. However, the accuracy of the method using cell-ID or enhanced observed time difference (E-OTD) is generally low (in the range of 50-200 m), depending on the cell size. Generally speaking, the accuracy is higher in densely covered areas (e.g, urban places) and much lower in rural environments [CS98].

2.2.3 Vision Based Approaches for Localization Systems

Vision based approaches use images from cameras in order to detect and track mobile agents. A video camera projects a tridimensional scene onto a bidimensional image plane, where each pixel loses the depth component. One could wonder whether, in the case we are only interested in the position of a mobile agents on the floor, the problem could be solved using a ceiling mounted camera; the problem here is that in the vast majority of the cases the field of view of a camera would allow to monitor only a very little section of the floor, so cameras are usually placed in a convenient position allowing to maximize the size of the monitored area.

[6] http://www.snaptrack.com
[7] http://www.ensco.com/products-services/gps-denied-geolocation-navigation.htm

The first problem we need to solve in order to define a vision based location sensing system is to choice a way to recover the third dimension. There exists a set of different techniques to do this:

- Systems relying on a single camera usually exploit a **3d model of human body** in order to deal with the lack of depth. This approach was used into the first vision-based systems and continues leveraging a certain interest in the research field. For example [ZNW08] uses a human model made up by a limited number of ellipsoids. For each agent the model can be configured using two parameters, namely the *size*, which models height and scaling of the human, and the *thickness*, which captures the extra scaling in the horizontal direction.

- Systems relying on multiple *time synchronized* cameras can exploit *n-View* geometry [HZ04]. In particular the so called **Stereo Vision** localization systems [BIL+07], [BI09], [Har04] exploit 2-View geometry. A continuously up-to-date comparison of techniques for stereo correspondence is available at `http://vision.middlebury.edu/stereo/` following the taxonomy defined in [SS02] and [SCD+06]. High accuracy stereo map can be computed using **structured light** [SS03]. In particular the recent advent of Kinect, which exploits IR structured light (the commercial system Light Coding produced by PrimeSense) to compute a depth map, opens new scenarios for vision based localization systems.

- Some kind of systems, especially robots, exploits **3d laser and ultrasound** scanner in order to obtain the third dimension of the scene. Merging data collected by the camera with measurements obtained by the laser is called **sensor fusion**.

Available vision-based systems for localization are mainly **marker-less**. Marker-less systems (see section 2.2.1) has to deal with a set of problems that are easier to solve with a marker-based systems, namely *background modeling*, *tracking* and *recognition*.

Background Modeling and Foreground Segmentation

Background modeling presents, in any kind of marker-less systems, a major difficulty in both long and short term *background dinamicity*. For *long term dinamicity* we refer to the fact that the positions of the objects in the scene as well as the scene itself could change over time (a chair is a good example of this, but also long term *illumination changing of the scene*), so background modeling techniques should contain a bootstrap phase where the initial model is defined but has to include an online mechanism to update the model while the scene is changing [CFBM10]. For *short term dinamicity* we refer to all those phenomena caused by periodic activity of the background, namely tree leaves moving, curtains, ventilators, screens changing their content and so on. So another important component of the background modeling is the model itself which is in charge of allowing the recognition of short term dinamicity..

Under the term Mixture of Gaussians (MOGs) there is a lot techniques based on the fact that each pixel in the image plane can be modeled using a set of gaussians distributions which clusters RGB data based on their distance in the color space. A subclass of MOGs method is made up by Time Adaptive Per-pixel Mixture of Gaussians (TAPPMOGs). While the basic concept is really simple implementation of TAPPMOGs may vary. For example [Har04] introduce a stereo vision system which makes use of the TAPPMOG described in [HGW01]. Here time adaptivity is reached applying an algorithm derived from Expectation-Maximization (EM). The TAPPMOG is applied not only to the RGB scene but also to the depth map. TAPPMOG is also used in [BIMR09] and [BIL+07] with time-adaptivity obtained creating an activity image based on the integration over time of Sobel filter for border detection. Simple MOGs are also used (without time adaptivity) in traffic analysis [BI09].

When a new frame is acquired, it is immediately compared to the background model in order to extract the scene foreground. This operation is conceptually simple and can be described as a subtraction. Nevertheless in the case of vision based systems, the foreground extraction algorithm has to detect video artifacts namely *shadows*, *sudden illumination changes* and *light hot-spots*. This issue can be solved using different strategies which can be based on the chosen background models or on the choice of different color spaces.

The extracted foreground has to be **segmented** in order to obtain a set of detected agents. This operation could be complex especially in the case of crowded scene. Here the simple extraction of blobs can create problems due to partial occlusions between agents. Recently solutions to crowding have been proposed in [BIMR09] and [ZNW08].

Tracking

In marker-based system, tracking is a simple operation given that each single marker may contain an ID that allows to retrieve a trajectory in space for the correspondent agent, so in the vast majority of the cases these trajectories have only to be smoothed detecting outliers and noise in measurement (Bayesian filters, such as Kalman Filters and Particle Filters, are used usually to do it). A Marker-less system has to associate to each agent in the scene a state including a set of features; that is marker-less systems maintains a set of tracked agents whose states are composed by features which are updated frame by frame and which are used to match agents extracted in the foreground of the current perception frame (which correspond to a video frame in vision-based systems). The problem is made even more complex by short term and long term occlusions that are frequent in especially in vision-based systems.

An extended class of vision-based (mostly composed by stereo-vision systems) systems execute the tracking over a simulated ceiling view of the environment [Har04], [BIL+07], [BIM+09]. In particular [BIM+09] uses different kind of trackers depending on how much the scene is crowded. For lowly crowded scenes a Single-Hypothesis Tracker (SHT) together with color features is used for tracking an agent. For highly crowded scenes a Multiple-Hypothesis Tracker (MHT) is used for each agent.

2.2.4 Identity Recognition

Identity Recognition is a biometric task implying the association of an identity to subject present in a monitored environment. The association of an Identity Recognition system to a Localization systems creates a powerful tools for the definition of intelligent services.

In Marker-based systems, identity recognition can be easily addressed knowing the association between a tag ID and the agent who wear the tag. In marker-less systems the task may be much more complicated; a particular source of problem is the fact that, for certain kind of applications (such as security), agents are not always truthful and so they will probably behaves against the identity recognition system.

In vision-based systems identity recognition is usually addressed using **face recognition** or more generally **physiological biometric**. An emerging research field called **behavioral biometric** tries to recognize humans base on physical or behavioral cues. Currently the most promising example of behavioral biometric is human gait [SPL+05].

Making agents, especially humans, unaware of the ongoing identification process in vision based systems obtaining a good performance is difficult due mostly to unconstrained lighting conditions and face orientations. So surveillance by face recognition systems has a low user satisfaction level [JL05].

2.3 Related Works: Activity Recognition

A good introduction to machine recognition of human activities (or simply activity recognition) is given in [TCSU08]. Here a distinction is done between the terms “action” and “activity”. In particular an action is defined as a “simple motion pattern usually executed by a single person and typically lasting for short duration of time” while an activity is defined as a “complex sequence of actions performed by a set of humans who could interact with each other in a constrained manner”.

The kind of activity recognition we want to perform is constrained by the kind of systems we can rely on. In particular activity recognition can be combined with a localization system or not.

Suppose we have a localization systems which returns the trajectories of a set of agents. Trajectory analysis have been widely used for activity recognition. A survey of these techniques can be found in [CS95]. More recently, research effort have been pushed to the definition of more refined techniques for action and activity recognition.

Now suppose we have access to a video sequence of the agent we want to derive the action or activity. Approaches for modeling actions can be divided mainly into three classes:

- *non-parametric approaches* are based on the accumulation of set of features of each single agent frame by frame. All these observations are then combined to represent the time evolution of an action event loosing in the representation the time dimension.
- *volumetric approaches* uses the video to obtain an, even simplified, 3d representation of the agent. This kind of approach is used for example in [PI08] and [SFC+11].
- *parametric approaches* specifically impose a model on temporal dynamics of the action. Such kind of models are based on Hidden Markov Models (HMM), Linear Dynamical Systems and Nonlinear Dynamical systems.

Modeling an activity is, in a certain sense, an higher level task with respect to action modeling and recognition. The reason is that, an activity may involve several subjects, each performing some actions, potentially interacting one with each other and together with the environment. So activity recognition can be seen as a process that take as input agents actions and contextual information and uses some sort of hardwired or learned knowledge to return the ongoing activity. Starting from this description models for activities can be divided into three classes:

- *graphical models* includes Dynamic Bayesian Networks (DBNs) and Petri Nets. Petri Nets are especially useful in order to model sequencing, concurrency, synchronization and resource sharing.
- *syntactic approaches* uses grammars to express the structure of a process using a set of production rules. In particular production rules specify how sentences (activities) can be constructed from words (activity primitives) and how recognize if a sentence (video) conforms to the rules.
- *logic-based approaches* uses formal logic rules to describe domain knowledge in order to describe activities. For example [NM08] introduce the use of description logics for scene interpretation.

It is worth noting how knowledge about environment could be very important for activity recognition. For example a trajectory pattern could have more sense if we know what symbolic location are near to the trajectory. An example use of contextual information (optionally obtained online) in order to improve an intelligent task have been shown in [CIN+09].

2.4 Research Proposal

My first Ph.D. year was devoted to the study of techniques for location and pose sensing and activity recognition. I've been particular interested in their applications for security monitoring and ambient automation.

My initial point will be the study of indoor localizations and activity recognition systems as low level service providers to be used as building blocks for higher level services. Final high level application influences the choice of a particular technique for localization sensing and activity recognition. Also, because of the cost of deployment of a specific system a trade-off between performance and cost with respect to the requirements of the final application has to be found.

Performance evaluation of localization and activity recognition systems has to be assessed clearly providing an universal test bed. During my first Ph.D. year I worked to the first version of a performance evaluation platform for indoor localization systems based on cheap hardware [LMA11]. I will extend and validate it using as a test bed a prototype of accessible house available at Fondazione Santa Lucia (a world wide known institute for motory rehabilitation). The extensions to the platform will include methods for assess the suitability of a particular localization system for an activity recognition task; for example some activity recognition tasks do not behave well if the underneath localization system is not able to manage occlusions. As stated in [TCSU08] while a set of test beds do exist for basic activity recognition tasks but no one has been presented for more complex activities.

Localization and activity recognition techniques evaluated will be used as building blocks for intelligent services to be used in the context of the *GreenerBuildings* European project. GreenerBuildings' aim is to develop an integrated solution for energy-aware adaptation of public buildings investigating the use of different technologies including self-powered sensors and actuators, occupant activity and behaviour inference, and an embedded software for coordinating thousands of smart objects with the goals of energy saving and user support. Additionally the project will be a chance for experimenting hybrid solutions for localization which merge techniques from both marker-less and marker-based world.

To summarize, the goals I would like to reach during my Ph.D. are strongly inherent to intelligent services for indoor and local outdoor:

- performance evaluation of localization and activity recognition systems;
- using knowledge based techniques for activity recognition and decision processes for building automation;
- devising hybrid solution for localization of human agents.

Bibliography

[BI09] D. D. Bloisi and L. Iocchi. ARGOS - a video surveillance system for boat traffic monitoring in venice. *International Journal of Pattern Recognition and Artificial Intelligence*, 23(7):1477–1502, 2009.

[BIL+07] S. Bahadori, L. Iocchi, G. Leone, D. Nardi, and L. Scozzafava. Real-time people localization and tracking through fixed stereo vision. *Applied Intelligence*, 26:83–97, 2007. 10.1007/s10489-006-0013-3.

[BIM+09] D. D. Bloisi, L. Iocchi, L. Marchetti, D. N. Monekosso, and P. Remagnino. An adaptive tracker for assisted living. In *IEEE Conf. on Advanced Video and Signal Based Surveillance*, pages 164–169, Los Alamitos, CA, USA, 2009. IEEE Computer Society.

[BIMR09] D. D. Bloisi, L. Iocchi, D. N. Monekosso, and P. Remagnino. A novel segmentation method for crowded scenes. In *Proc. of 4th Int. Conf. on Computer Vision Theory and Applications (VISAPP-2009)*, pages 484–489, 2009.

[BP00] P. Bahl and V.N. Padmanabhan. Radar: an in-building rf-based user location and tracking system. In *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 2, pages 775 –784 vol.2, 2000.

[CFBM10] M. Cristani, M. Farenzena, D. Bloisi, and V. Murino. Background subtraction for automated multisensor surveillance: A comprehensive review. *EURASIP Journal on Advances in Signal Processing*, 2010, 2010.

[CIN+09] D. Calisi, L. Iocchi, D. Nardi, G. Randelli, and V.A. Ziparo. Improving search and rescue using contextual information. *Advanced Robotics*, 23(9):1199–1216, 2009.

[CS95] C. Cedras and M. Shah. Motion-based recognition a survey. *Image and Vision Computing*, 13(2):129 – 155, 1995.

[CS98] J.J. Caffery and G.L. Stuber. Overview of radiolocation in cdma cellular systems. *Communications Magazine, IEEE*, 36(4):38 –45, apr 1998.

[Har04] M. Harville. Stereo person tracking with adaptive plan-view templates of height and occupancy statistics. *Image and Vision Computing*, 22(2):127 – 142, 2004. Statistical Methods in Video Processing.

[HBC+05] A. Hampapur, L. Brown, J. Connell, A. Ekin, N. Haas, M. Lu, H. Merkl, and S. Pankanti. Smart video surveillance: exploring the concept of multiscale spatiotemporal tracking. *Signal Processing Magazine, IEEE*, 22(2):38 – 51, march 2005.

[HGW01] M. Harville, G. Gordon, and J. Woodfill. Foreground segmentation using adaptive mixture models in color and depth. In *Detection and Recognition of Events in Video, 2001. Proceedings. IEEE Workshop on*, pages 3 –11, 2001.

[HHS+02] Andy Harter, Andy Hopper, Pete Steggles, Andy Ward, and Paul Webster. The anatomy of a context-aware application. *Wireless Networks*, 8:187–197, 2002. 10.1023/A:1013767926256.

[HWB00] Jeffrey Hightower, Roy Want, and Gaetano Borriello. SpotON: An indoor 3d location sensing technology based on RF signal strength. UW CSE 00-02-02, University of Washington, Department of Computer Science and Engineering, Seattle, WA, February 2000.

[HZ04] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004.

[JL05] Anil K. Jain and Stan Z. Li. *Handbook of Face Recognition*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.

[JLY09] C.S. Jensen, Hua Lu, and Bin Yang. Graph model based indoor tracking. In *Mobile Data Management: Systems, Services and Middleware, 2009. MDM '09. Tenth International Conference on*, pages 122 –131, may 2009.

[JLY10] Christian S. Jensen, Hua Lu, and Bin Yang. Indoor - a new data management frontier. *IEEE Data Eng. Bull.*, 33(2):12–17, 2010.

[LDG09] Zheng Li, W. Dehaene, and G. Gielen. A 3-tier uwb-based indoor localization system for ultra-low-power sensor networks. *Wireless Communications, IEEE Transactions on*, 8(6):2813 –2818, june 2009.

[LMA11] F. Leotta, M. Mecella, and F. Aloise. Pericles: A performance evaluation platform for indoor localization systems. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness*, ISA '11, 2011. To Appear.

[NLLP04] Lionel M. Ni, Yunhao Liu, Yiu Cho Lau, and Abhishek P. Patil. Landmarc: Indoor location sensing using active rfid. *Wireless Networks*, 10:701–710, 2004. 10.1023/B:WINE.0000044029.06344.dd.

[NM08] B. Neumann and R. Moller. On scene interpretation with description logics. *Image and Vision Computing*, 26(1):82 – 101, 2008. Cognitive Vision-Special Issue.

[PI08] S. Pellegrini and L. Iocchi. Human posture tracking and classification through stereo vision and 3d model matching. *J. Image Video Process.*, 2008:7:1–7:12, January 2008.

[RMT+02] T. Roos, P. Myllymaki, H. Tirri, P. Misikangas, and J. Sievanen. A probabilistic approach to wlan user location estimation. *International Journal of Wireless Information Networks*, 9:155–164, 2002. 10.1023/A:1016003126882.

[SCD+06] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1*, pages 519–528, Washington, DC, USA, 2006. IEEE Computer Society.

[SFC+11] 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 single depth images. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 1297 –1304, june 2011.

[SPL+05] S. Sarkar, P.J. Phillips, Z. Liu, I.R. Vega, P. Grother, and K.W. Bowyer. The humanid gait challenge problem: data sets, performance, and analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(2):162 –177, feb. 2005.

[SS02] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47:7–42, 2002. 10.1023/A:1014573219977.

[SS03] D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 1, pages I–195 – I–202 vol.1, june 2003.

[TCSU08] P. Turaga, R. Chellappa, V.S. Subrahmanian, and O. Udrea. Machine recognition of human activities: A survey. *Circuits and Systems for Video Technology, IEEE Transactions on*, 18(11):1473 –1488, nov. 2008.

[TGLP08] R. Tesoriero, J. Gallud, M. Lozano, and V. Penichet. Using active and passive rfid technology to support indoor location-aware systems. *Consumer Electronics, IEEE Transactions on*, 54(2):578 –583, may 2008.

[VRB+10] F Viani, P Rocca, M Benedetti, G Oliveri, and A Massa. Electromagnetic passive localization and tracking of moving targets in a wsn-infrastructured environment. *Inverse Problems*, 26(7):074003, 2010.

[WHFaG92] Roy Want, Andy Hopper, Veronica Falcão, and Jonathan Gibbons. The active badge location system. *ACM Trans. Inf. Syst.*, 10:91–102, January 1992.

[ZNW08] Tao Zhao, Ram Nevatia, and Bo Wu. Segmentation and tracking of multiple humans in crowded environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30:1198–1211, 2008.