Matthias Kranz, Eckehard Steinbach, Andreas Möller, Luis Roalter,
Michael Eichhorn, Florian Schweiger (Editors)

ADVANCES IN MEDIA TECHNOLOGY

MOBILE INTERACTION AND MOBILE SERVICES

February 10, 2011

Technische Universität München
Institute for Media Technology/
Distributed Multimodal Information Processing Group

MOBILE INTERACTION AND MOBILE SERVICES

Today's computer systems are becoming more and more ubiquitous, and the increasing processing power and capabilities of modern mobile devices enable new applications and services. Integrated sensors provide new ways of how users can interact with their environment and are a basis for a wide field of context-aware services.

In this seminar, various aspects of mobile interaction and mobile services were investigated. Being a key requirement for many mobile services, approaches for the estimation of location under different conditions were investigated, and current research in the area of location-based services and context-awareness was discussed. The second focus of the seminar was set on interaction techniques between mobile devices, context, and the physical environment, and in what way feedback information is presented to users, with regard to the limitations, but also the potential of sensor-equipped mobile devices.

Each of the single chapters addresses a specific topic and presents fundamental research trends and developments. This report is targeted at electrical engineers, computer scientists and anyone interested in technology and realization of mobile services and corresponding new ways of interaction.

The seminar has jointly been organized by the Institute for Media Technology and the Distributed Multimodal Information Processing Group at Technische Universität München.

This report contains the following scientific seminar theses:

1. Positioning Techniques beyond GPS
2. Indoor Navigation
3. Indoor Navigation 2
4. Activity Recognition
5. Ubiquitous Campus
6. Interaction with Mobile Navigation Systems
7. Mobile Interaction with the Physical World
8. A Survey of Interaction Techniques for Public Displays
9. Mobile Authentication and Security

The website with the electronic version of the proceedings and additional material and resources can be found at `http://www.lmt.ei.tum.de/courses/hsmt/hsmt.2010_WS.html`

We thank all students for the great work and their participation!

Munich, February 2011

Prof. Dr. Matthias Kranz
Prof. Dr. Eckehard Steinbach
Andreas Möller
Luis Roalter
Michael Eichhorn
Florian Schweiger
(Editors)

TABLE OF CONTENTS

Positioning Techniques beyond GPS

Zinan Shi
zinan.shi@mytum.de

February 10, 2011

Abstract

With the rapid development of communication and information technology, our demand of wireless localization is increasing day by day. This paper gives an overview of the existing wireless indoor positioning techniques. Several typical indoor positioning systems and the techniques they have used are described and analyzed. The possible trend of positioning system research are presented.

1 Introduction

Ubiquitous computing is a major trend of the next generation information technology. The idea of ubiquitous computing is that, computing (information processing) should be integrated into our daily life. That means all things equipped with embedded processor can connect with networks to compute and exchange information. As an important requirement of ubiquitous computing system, indoor localization is obtaining more and more attentions. There are many applications depending on indoor localization technologies in our daily life. For example the location of the stored things could be detected in a warehouse, the location of medical personnel and equipment in a hospital could also be detected. The indoor location application at certain emergency service departments, such as police, military and firehouse is particularly important.

Global Positioning System (GPS) is a world-wide used outdoor localization system nowadays. GPS system uses four satellites as its reference points. Three satellites should be enough for 2-D positioning, but one more satellite is required, due to clock accuracy. The trilateration algorithm is used to calculate the location of the target object. This system can achieve an accuracy at 10m with 50% probability. But GPS can't be used for indoor localization due to the satellite signal attenuation. The GPS satellite signal is too weak to penetrate building.

In the second section the structure of a typical positioning system and some approaches to determine location will be introduced. Some typical indoor localization systems, the techniques they have used, their advantages and disadvantages will be described in the third section.

2 Positioning System Structure and Positioning Algorithms

2.1 Positioning System Structure

A typical positioning system (see Fig. 1) is composed of the location sensing part, the location computation part (use corresponding positioning algorithms) and the display part. The process of collect the information from the location of a target object is called location sensing, this part can be continue subdivided into two parts: a signal transmitter and a signal receiver. In the transmitter some source signals such as WiFi, GSM, DECT, Bluetooth, ultra sound, or IR (infrared) are used to sense the target object. The receiver collects the feed-back values, such as RSS (Received Signal Strength), AOA (Angle of Arrival), TOF (Time of Flight) or TDOA (Time Difference of Arrival), for the location computation part. With corresponding algorithms the location of target object is computed and displayed.

2.2 Positioning Algorithms

Various approaches are used to determine location. They all have own advantages and disadvantages. In this section several fundamental techniques[1] and general concepts for building localization systems, such as trilateration, proximity, hyperbolic lateration, triangulation, fingerprinting and dead reckoning, will be described.

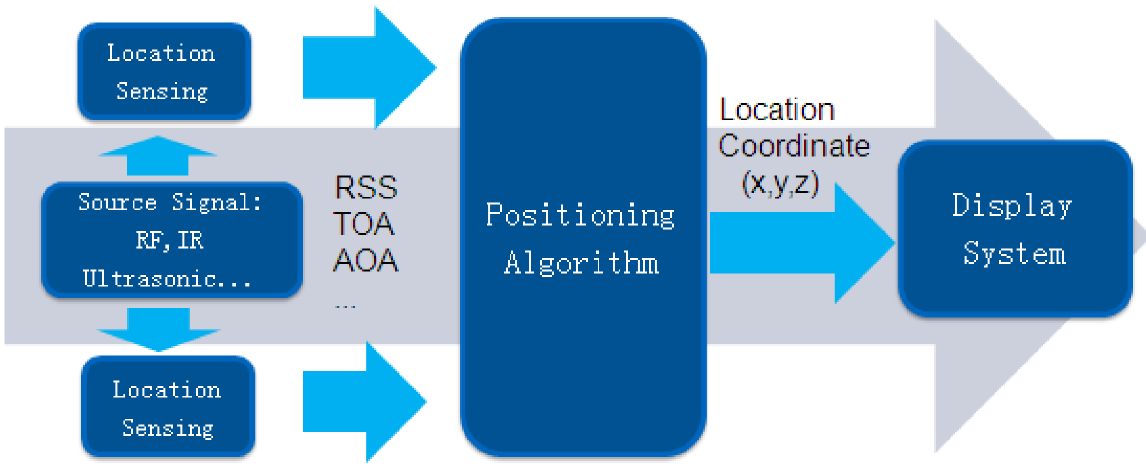


Figure 1: Basic architecture of the positioning system. A typical positioning system is composed of the location sensing part, the location computation part and the display part.
RF: Radio Frequency; IR: Infrared; RSS: Received Signal Strength; TOA: Time of Arrival;
AOA: Angle of Arrival;

2.2.1 Trilateration

Trilateration is a very important localization algorithm and widely used in many positioning systems. As long as we know the distance between the target object and several reference points, the location of the target object can be calculated with the trilateration localization algorithm. There are two common approaches to estimate the distance between a target object and the reference point: Using TOF (Time of Flight) or RSS (Received Signal Strength).

We know the speed of sound and the speed of light in propagation process (344 m/s in air for sound and 299,792,458 m/s for light). If we already know the time that the sensing signal needs to travel between the target object and the reference points (TOF), the distance between them can be calculated through the product of the measured TOF and the speed of source signal. Precise synchronization clock between the target object and the reference point is required for the measurement of the TOF. This is especially important as estimating the TOF of a radio or light signal because small mistakes of the clocks will result large measurement errors.

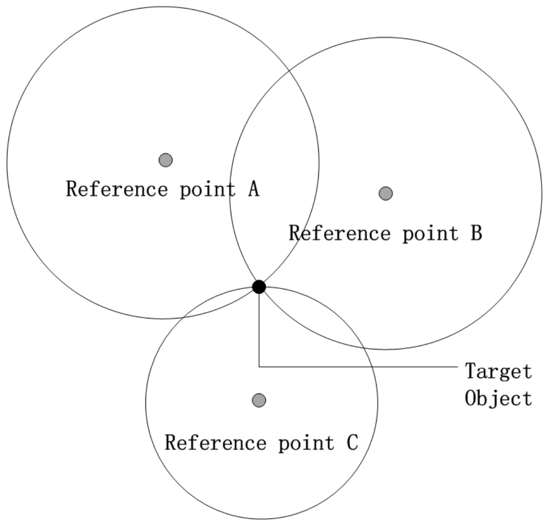
Using the measured decrease of RSS (received signal strength) is another approach to estimate the distance between a target object and a reference point. The attenuation model of the used source signal is required for this calculation. The signal attenuation model describes the signal propagation process. For example, the signal attenuation model of radio frequency (RF) [1] states that the strength of a radio signal decreases by a factor of $1/r^2$ (r is the distance from the radio source). If we know the attenuation model of a signal and the measured RSS (received signal strength) at a certain location, we can obtain the distance between the signal source and the target location.

But there is a technical challenge to describe the signal attenuation during the propagation with a precise model due to multipath effect, low probability for availability of line-of-sight (LOS) and specific location parameters such as floor layout, moving objects. There is no perfect model for the indoor radio multipath characteristic so far[2].

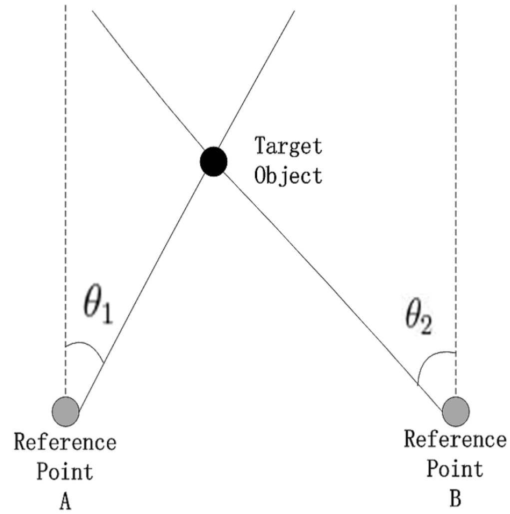
Fig. 2 (a). shows an example of trilateration localization algorithms. The gray dots are the reference points. Now we have estimated the distance between a target object and a reference point. We can suppose that, the reference point is the center of a circle with a radius equal to the estimated distance to the target object. Then we can get an infinite number of possible locations of the target object on the border of the circle. With the distance between the target object and two reference points, we can define two circles. At the intersections of the two circles we can obtain two possible locations of the target object. By a similar way we can get the unique location of the target object, if we know the distance between the target object and three reference points in 2-D scenario.

2.2.2 Proximity

Proximity algorithm describes a scenario, in which several Base-stations or APs (access points) are deployed to estimate the location of the target object. If a target object communicates with a basis station or is detected by an AP,



(a) An example of Trilateration. The gray dots are reference points. The black dot represents the target object. The location of the target object is uniquely defined by the three reference points.



(b) An example of Triangulation. The gray dots are reference points. The black dot represents the location of the target object. θ_1 and θ_2 are the AOA (angle of arrival) from the target object to the two reference points

Figure 2: Trilateration and Triangulation algorithms, which are usually used to estimate the location of the target object.

this means the target object is in the range, that the base-station (AP) signal can arrive. When more than one AP has detected the target object, we can know a more precise location of the target object in the overlapped area, which the APs can cover. And it is possible to get more accurate location estimation by using the algorithm of weighted average of the reference points' positions, which will be introduced in the Fingerprinting section.

2.2.3 Hyperbolic Lateration

Unlike the Trilateration using the TOF to estimate the distance between a target object and reference points. Hyperbolic Lateration uses the TDOA (time difference between the signal arrival times) between the target object and several reference points to estimate the location of the target object.

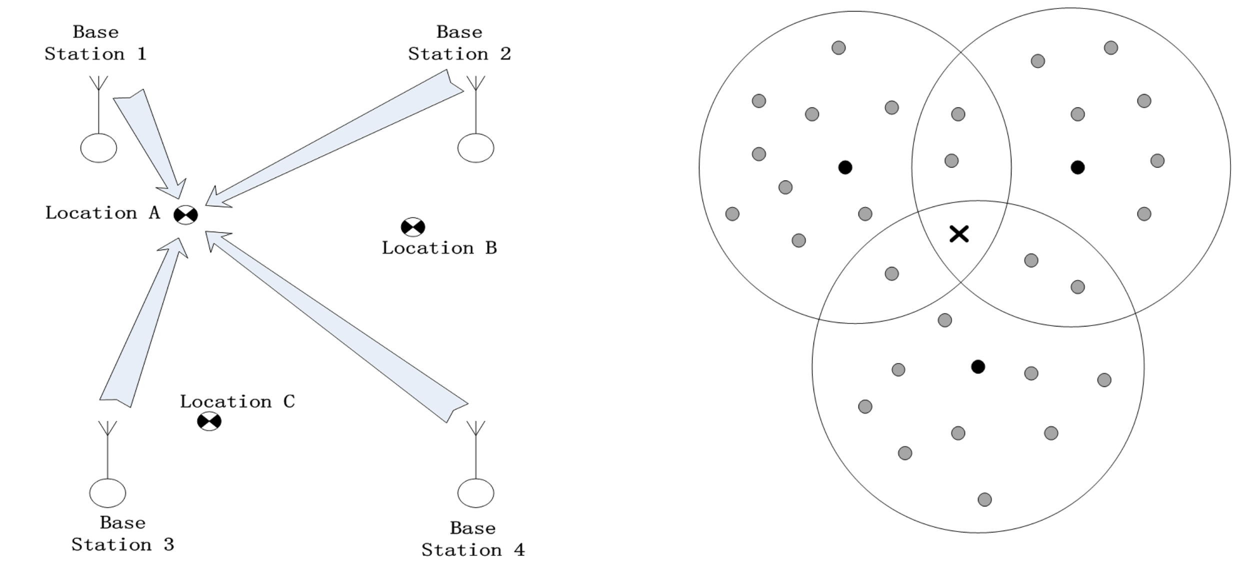
The principle of the Hyperbolic Lateration: the TDOA between a target object and two different reference points defines the possible positions of the target object along a hyperbolic line (the two reference point serve as foci of the hyperbolic line). If we know another reference point and the corresponding TDOA, the other hyperbolic line can be defined. We can obtain the unique location of the target object at the intersection of the two hyperbolic lines.

2.2.4 Triangulation

As already discussed above, TOF (Time of Flight) is used in the Trilateration algorithms to estimate the distance between target object and reference points. But expensive and accurate synchronization clock is required for the precise measurement of the TOF. This increases the difficulty to deploy such systems.

So there is another idea that uses the AOA (angle of arrival) from a target object to two reference points to estimate the location of the target object. We can easily see that, there's no time synchronization requirement between measuring units in this algorithms. The AOA can be measured with directional antenna or an array of antennae.

Fig. 2 (b). shows an example of Triangulation. The gray dots A and B are the reference points. The AOA (angle of arrival) from the target object (black dot) to point A and B are θ_1 and θ_2, which are required in the Triangulation. With point A and the angle θ_1 we can draw an angle's directional line and we do the same at point B. The unique position of target object can be obtained at the intersection of the two lines.



(a) An example of Fingerprinting Algorithms. There are four Base-stations which emit RF signal. The RSS (reiceived signal strength) difference in location A, B, C define their own fingerprinting.

(b) LANDMARC Positioning System. The black dots are RFID (radio frequency identification) readers, the gray dots are reference tags and the cross is the target object.

Figure 3: Fingerprinting Algorithms,which is widely used in positioning systems.

Another advantage of Triangulation algorithms is that, there are only 2 reference points required to estimate the location of the target object. In contrast, Trilateration and Hyperbolic Lateration require 3 reference points for their calculation in the 2-D scenario.

In the indoor environment if the signal travels in a direct line, it can't reach some particular area. But this reality is broken due to numerous reflecting surfaces in the indoor environment. In the signal propagation process, the signal may be reflected, refracted or diffracted, thus signal changes its propagation direction. Multipath effect can influence the accuracy of angle measurement. This is also a big challenge for the positioning systems which use triangulation algorithms.

2.2.5 Fingerprinting

Every person has his unique fingerprint and it can be used in personal identification. Similarly in an indoor environment, the RSS (received signal strength) changes in different locations due to the variability of the RF (radio frequency) signal. But the RSS of the RF signal stays always the same in a fixed location over time. These two properties make it possible to use the fingerprinting algorithm to determine the location of the target object. RSS based location fingerprinting is a very important algorithms which has been used in many location systems[3].

A simple example is shown in Fig. 3(a). We see four Base-stations which emit RF signal and three locations which should be investigated in this scene. For example we receive a signal with 80% strength from station1 at location A, 40% from station2, 50% from station3, only 20% from station4. We write the measured RSS in a vector form: (0.8, 0.4, 0.5, 0.2). This RSS vector is the fingerprint of location A. Location B has a fingerprint (0.2, 0.7, 0.2, 0.2) and (0.1, 0.1, 0.2, 0.9) is measured at location C. If we get a RSS like (0.1, 0.1, 0.2, 0.8) at one point, through comparison with the fingerprint from A B C, we know the position is near location C. That means, this positioning technology uses a pattern matching technique to estimate the location of the target object.

A typical fingerprinting-based localization system is composed of two processes: the off-line phase and the on-line phase. In the off-line phase we build a signal map for our indoor-environment, the fingerprints from different locations are collected and stored in a database. In the online phase the RSS from the location of the target object is measured and transmitted to the server. The serve estimate the location of the target object by matching online measurements with the closest location fingerprints in the database.

We compare the similarity between the current measured RSS and the fingerprints in database (vectors) by calculating the Euclidean distance between them. For example, the fingerprint of one reference point in the database is $(R_1, ..., R_n)$. And the fingerprint at the location of target object is $(C_1....C_n)$. Then the Euclidean distance E between them should be calculated as:

$$E = \sqrt{(R_1 - C_1)^2 + (R_2 - C_2)^2.... + (R_n - C_n)^2}$$

The (NN) nearest neighbor algorithms finds out the referent point with the minimal E value in the database. There are also some other algorithms such as K-NN (k-nearest-neighbor), neutral networks, support vector machine (SVM), and smallest M-vertex polygon (SMP)[2].

Received signal strength can be affected by diffraction, reflection, and scattering during the propagations process in the indoor environment and this is a main challenge for such techniques based on location fingerprinting. We will discuss about this further in the next section with RADAR and LANDMARC positioning system.

2.2.6 Dead Reckoning

Physical knowledge tells us, if we know the average speed and the direction of a moving car, we can estimate its position in the next couple of seconds. The idea of dead reckoning is based on the estimation of the average speed, direction of the movement at the last location to calculate the current location of a target object. The quality of estimating the average speed and direction of the movement affects the accuracy of dead reckoning.

Some location systems can be complemented by the dead reckoning technology when the current system is temporary out of service. For example when a car enters a tunnel and loses the signal from GPS satellites. In this time quantum dead reckoning can serve as a temporary substitute positioning system.

3 Selected Localization Systems

In this section several typical indoor positioning systems and the techniques they have used will be discussed.

3.1 Active Badge

Pulse-width modulated infrared (IR) is used as the source signal to detect the location of target objects. Based on this idea, the first indoor automated location system Active Badge[4] is developed by Olivetti Research Ltd. (ORL) in Cambridge.

The Active Badges are designed to emit unique code, which contains the information of target object by using periodical IR signals (every 15 seconds). Testers wear the active badges in order to transmit their unique ID information to the sensors. The sensors are deployed throughout the test space (for example, one sensor per room, more sensors in big rooms, such as conference rooms) and the sensors are through a network with another connected.

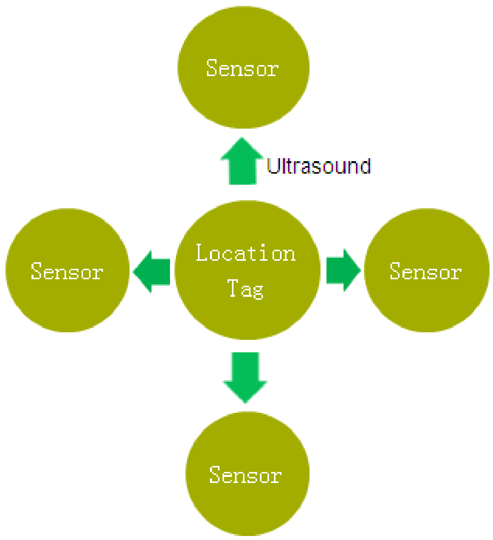
The Proximity algorithm is used in this system to estimate the position of the testers. The sensors are deployed in fixed locations and the IR signal can't travel throughout the wall. Thus if the target object is detected by a sensor, we can know that the target object is in the same room with this sensor. The information with the current locations of the testers is stored in a central server.

The accuracy of this system stays at room-level. The LOS (line of sight) is important for the IR signal. Thus the density and the placement strategy of the sensors affect the accuracy of this positioning system.

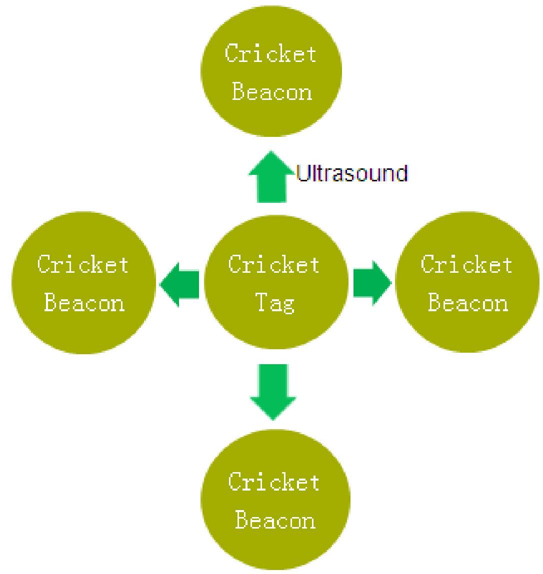
3.2 Active Bat

Ultrasonic can also be used as the source signal in positioning systems to detect the location of target objects. The Active Bat system [5] is one example.

The Active Bat system is composed of ultrasound receivers and location tags. Fig.4(a) is a simple example of this system. Receivers are deployed throughout the test environment and the positions of the receivers are fixed. The receivers serve in this system as reference points. They can detect the ultrasonic signal, which is emitted by the active location tag. And the TOF between the tag and reference point should be measured. As already related in the second section, a synchronization clock is very important for the TOF measurement between the signal emitter and receiver. In Active Bat system RF signal serves as the Synchronization Clock. The RF synchronization signal gives the location tag



(a) Active Bat system has an active structure.This means, the location tags (target object) actively emit source signal to sensors (reference points).



(b) Cricket system has a passive structure.This means, the location tags (target object) passively receive sensing signal from Cricket beacons (reference points).

Figure 4: Active and Passive Architecture of Localization System

a hit to transmit its ultrasonic pulse, at the same time it also gives the receivers a starting point for timing the received ultrasonic signal.

With multiplication between the speed of sound and the measured TOF, the distances from the location tag (target object) to each receiver (reference point) are calculated. Then the position of the location tag can be determined by using trilateration algorithms.

If there are many target objects (location tags) in our positioning system which simultaneously transmit their Ultrasonic signal to the receivers, the IDs of the location tags are distinguished by the receivers through the different ultrasound bandwidth they have used. The Active Bat system has an active system structure as shown in Fig.4(a), thus the location tags must coordinate the bandwidth of the ultrasound signal they have used, so as not to interfere with each other. The difficulty becomes larger when the density of location tags increases. The LOS (line-of-sight) requirement must also be considered during the system deployment.

The Active Bat system supports 75 tags, which can be tracked in a 1000 square meters space with 720 receivers. This system achieves a localization accuracy of 90% at 3 centimeters.

3.3 Cricket

As already discussed above, the Active Bat or Active Badge have a centralized and active system architecture. This means the target objects emit source signal (e.g. IR, Ultrasound), which carries their information actively to the receivers and the location information is calculated and stored in a central server. Because the central server knows the location information from all the target objects, this leads to a privacy problem.

In contrast, the Cricket location system [6] is based on a passive system architecture to compute location information. Fig. 4(b).) is a simple example of this system. Each Cricket tag is composed of an RF transmitter, a microcontroller and a receiver for ultrasonic signals. Cricket beacons are deployed with fixed locations throughout the test environment. The Cricket beacons serve as signal emitters and reference points in this system. Ultrasound signals are emitted from the Cricket beacons to the Cricket tag, at the same time a RF signal, which serves as a synchronization clock and contains position information of Cricket beacons, is also transmitted to tell the Cricket tag the starting point for timing. The TOF between the Cricket beacons and target tag are measured and the distances between them are calculated. Trilateration is used to compute the location of target objects with these distance values and the Cricket beacon (reference point) position information.

The passive architecture works better as the density of location tags increase, because the system deployment is independent of the number of tracking tags in the environment. The privacy can be also guaranteed. But due to the LOS (line-of-sight) requirement a suitable and sufficient deployment of the Cricket beacons is still required.

3.4 RADAR

There is one thing in common from the above discussed systems: Some special hardware infrastructure must be installed for the localization system (e.g. Active Badges, Cricket beacons, Cricket tags). This increases the cost and complexity to deploy those positioning systems for our practical application.

Radar (proposed by Bahl and Padmanabhan) [7] is an indoor positioning system based on the existing hardware infrastructure (WiFi). Two kinds of approaches are used to estimate the target object location.

The first approach is using the RSSI (received signal strength indicator) to calculate the distance between target object and reference points. For example, in Fig. 3(a). we have four Base-stations (WiFi AP 1-4) and we receive at location A signal strength of 80% from AP 1. By using the wall attenuation factor (WAF) and floor attenuation factor (FAF) in indoor environment the distance between the AP1 and location A can be calculated. With the same approach the distance between A and AP2 or Ap3 is estimated. Then the location of point A is determined by using Trilateration algorithms.

The second approach is based on the fingerprinting algorithm. In the off-line phase the signal map is established. Some devices such as WiFi-enabled mobile phones, PDAs, or laptops are used as receiver to collect the RSS based fingerprints from several locations. A grid spacing of 1 meter is found to be a suitable value in many experiments. In the online-phase the RSS from target object position is compared to the fingerprints in database. The physical coordinate of the NN (nearest neighbor) point should be used to determine the target object position.

RADAR can achieve an accuracy of 3 meters in average and 90% of 6 meters. At least three APs are needed to be in range of the effective localization. One weak point of this system is that, the RSS from a location is affected by some environment changes (e.g. moving furniture, appliances, absorption). The in offline-phase built signal map can't accommodate the real-time changes. This has effect on the accuracy of this system.

3.5 LANDMARC

As related above, the environmental changes can't be accommodated in some positioning systems based on the fingerprinting technology, such as RADAR due to the in offline-phase completed signal map (fingerprint database). In contrast, the LANDMARC [8] positioning system which is developed by Michigan State University and Hong Kong University of Science & Technology together can easily sense the environmental changes, thus the location information is more accurate and reliable.

The LANDMARC uses the RFID (Radio Frequency Identification) technology for the localization and this system is composed of RFID readers, reference RFID tags whose positions are fixed. The readers read the data, which is emitted from the RFID tags. A communications network is constructed between them. Thus, the LANDMARC has a centralized system structure. Frequency ranges used for active RFID tags are UHF (868-915MHz) and microwave frequency (2.45GHz, 5.8GHz). A defined RF protocol is used for RF signal to transmit information between the readers and the tags. And active RFID tags are used in the LANDMARC system. This means the active tags emit continuously RF signal which contains their ID information to the RFID readers.

LANDMARC introduces the concept of reference tag to reduce the number of deployed RFID readers and guarantee the accuracy of localization.

A graphic description of LANDMARK is showed in Fig. 3(b). This system uses reference tags (the gray dots), whose positions are fixed to help location calibration. The idea of LANDMARC system is to find out the nearest neighbor reference tags for the tracking tag (cross). The readers (black dots) detect the signal strength from tracking tag and reference tags. Then the measured values (vectors) are compared and the Euclidian distances E between them are calculated. K-NN weighted value algorithm is used to determine the location of the tracking tag. This algorithm means: K reference points with the smallest Euclidean distances are selected for the target object position calculation. Between the K selected tags, the tags with smaller Euclidean distances have a higher weighted value. For example we can use a formula like this:

$$(X,Y) = \sum_{i=1}^{k} \frac{\frac{1}{E_i^2}}{[\sum_{i=1}^{k} \frac{1}{E_i^2}]}(X_i, Y_i, Z_i)$$, where (X_i, Y_i, Z_i) is the coordinate for reference point i.

In this section we see that, in contrast with RADAR system, the LANDMARC don't require the Signal-map (collected in the offline training phase). The current RSS changes due to diffraction, reflection, and scattering during the propagation process in indoor environments can be estimated by active tags. Instead of using the expensive RFID readers, cheaper active tags are used to estimate the environment changes. This increases the possibility for the deployment of such systems for practical application.

The LANDMARC system has an error probability of 50% at 1m and the maximum error is less than 2m.

3.6 GSM

GSM (Global System for Mobile Communications) can also be used in the positioning system to estimate the location of target object. In those systems the GSM bases stations are used as reference points. The key idea that makes accurate GSM-based indoor localization possible is using wide signal-strength fingerprints. The wide fingerprint contains the six strongest GSM cells and readings of up to 29 additional GSM channels, most of them are strong enough to be detected but too weak to be used for efficient communication.[9]

Like the RADAR system, RSS based fingerprints are collected from the indoor environment to build the signal map. And the weighted k-NN (k-nearest neighbor) algorithms is used to determine the location of the target object.

The results show that, the GSM based indoor localization system can achieve a median accuracy of 5 meter in large multifloor buildings.

3.6.1 Positioning Using DECT

Cordless phone system has a modern wireless communication infrastructure. Professor Kranz from Technische Universitaet Muenchen has in his paper [10] shown that, DECT (Digital Enhanced Cordless Telecommunications) can also be used to improve the accuracy and robustness of the existing localization schemes based on WLAN or GSM.

There are some advantages of the DECT based positioning system: There is no interference with other RF signals, because DECT has an exclusive frequency band. DECT can be used for both indoor and outdoor localization. Except the urban outdoor scenario, DECT has more base stations then WLAN, thus we expect a higher accuracy from the DECT based positioning system. In comparison with RSS, the stability of DECT signal is also very good.

Like the other localization systems, which use fingerprinting, the DECT based localization system collects Fingerprints at various locations and stores them in a look up table during the initial offline training phase. The Manhattan distance between fingerprints is used to evaluate the similarity. The location with smaller distance has the bigger similarity. The weighted K-NN (K-nearest neighbors) method is used to estimate the location of the target object. The locations of the K fingerprints with the smallest Manhattan distance to the target object are averaged (the same as already described in the LANDMARC section). The DECT system can achieve a fixed accuracy of 5 meters of indoor measurements.

At the end of this section, a summary about the available positioning systems is shown in Tab.1.

4 Conclusion and Future Trends

This paper gives an overview about the current existing indoor positioning systems and the techniques they have used. All the existing positioning systems have their own advantages and disadvantages. For example, the Cricket system can achieve a high accuracy (cm-level), but the LOS (line of sight) and a high density of beacons deployment are required due to the limited transmission ability of ultrasound. This increases the deployment cost. The RADAR system saves the cost of infrastructure deployment, but there is an additional training cost in the off-line phase. And the current environment changes can't be accommodated. So a positioning system which is with high accuracy, low-cost, low complexity (easy to deploy) should be found in the future. To develop such positioning system, a practical approach is to use the hybrid position algorithms. This means the existing positioning techniques are integrated into one system.

It is also be supposed, that we can in indoor environment and outdoor scenario use the same positioning system. Approaches, which join the indoor and outdoor positioning systems together, should be exploited. This technique can provide us a more efficient and robust positioning system.

A ultrasound based customer localization system is used by an electronic market called Best Buy from San Francisco. In this system radio direction-finding stations serve as reference points, which emit ultrasound signals. The ultrasound is received by the microphone of customer's mobile phone. The Triangulation algorithm is used to determine the location of customer. In this system the customers are guided to the expected products or to some products with special price. Such real-world application should be continuously exploited.

In some emergency scene (such as E-911 emergency responder application), it is very important to deploy the sensors as soon as possible. How to finish deploying wireless positioning system in a short time, how to deploy the sensors to improve the positioning accuracy, should also be considered.

References

[1] Krumm, J.: Ubiquitous Computing Fundamentals. Chapman & Hall/CRC (2009) 2.2, 2.2.1, 1

[2] Liu, H., Darabi, H., Banerjee, P., Liu, J.: Survey of wireless indoor positioning techniques and systems. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews **37**(6) (2007) 1067–1080 2.2.1, 2.2.5

[3] Kaemarungsi, K., Krishnamurthy, P.: Modeling of indoor positioning systems based on location fingerprinting. In: INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies. Volume 2., IEEE (2004) 1012–1022 2.2.5

[4] Want, R., Hopper, A., Falcao, V., Gibbons, J.: The active badge location system. ACM Transactions on Information Systems (TOIS) **10**(1) (1992) 91–102 3.1

[5] Harter, A., Hopper, A., Steggles, P., Ward, A., Webster, P.: The anatomy of a context-aware application. Wireless Networks **8**(2) (2002) 187–197 3.2

[6] Priyantha, N., Chakraborty, A., Balakrishnan, H.: The cricket location-support system. In: Proceedings of the 6th annual international conference on Mobile computing and networking, ACM (2000) 32–43 3.3

[7] Bahl, P., Padmanabhan, V.: 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., Ieee (2002) 775–784 3.4

[8] Ni, L., Liu, Y., Lau, Y., Patil, A.: LANDMARC: indoor location sensing using active RFID. Wireless Networks **10**(6) (2004) 701–710 3.5

[9] Otsason, V., Varshavsky, A., LaMarca, A., De Lara, E.: Accurate gsm indoor localization. UbiComp 2005: Ubiquitous Computing (2005) 141–158 3.6

[10] Kranz, M., Fischer, C., Schmidt, A.: A comparative study of DECT and WLAN signals for indoor localization. In: Pervasive Computing and Communications (PerCom), 2010 IEEE International Conference on, IEEE (2010) 235–243 3.6.1

Location System	Location Type	Resolution Accuracy	Infrastructure Requirements	Location Data Storage	Spectral Requirements	Location Systemtype
Active Badge	Symbolic Indoor	Room level	IR Sensors Customs tag	Central	IR	Custom active tagging
Active Bat	Absolute Indoor	3cm,90%	Ultrasonic Receiver Transmitter	Central	30kHz ultrasound 900MHz RF	Custom active tagging
Active Floor	Symbolic Indoor	1m,91%	Custom floor tiles	Central	Load sensor	Passive
Airbus	Symbolic Indoor	Room level,88%	Single sensor in HVAC	Central	Pressure sensor	Passive
Cricket	Absolute Indoor	3cm,90%	US receivers and transmitters	Local	30kHz ultrasound 900MHz RF	Custom active tagging
GPS	Absolute outdoor	10m,50%	GPS receivers	Local	1500MHz RF	Custom active tagging
PlaceLab (GSM)	Symbolic Indoor/outdoor	20m,90% 5m,50%	Existing GSM towers	Local	900-2000MHz RF	Custom active tagging
RADAR	Symbolic Indoor	6m,90% 2-3m,50%	3-5 WiFi APs	Local	2.4GHz RF	Custom active tagging
LANDMARC (RFID)	Symbolic Indoor	1m,50%	RFID readers and tags	Central	868-915MHz RF	Custom active tagging
DECT	Symbolic Indoor/Outdoor	Indoor 5m 20m 50%	DECT stations	Local	1880-1900MHz RF	Custom active tagging
Vision	Absolute Indoor/Outdoor	1m,50%-80%	Multiple cameras	Central	RF for wireless cameras	passive

Table 1: Overview of several Location Systems.[1]

Indoor Navigation

Simon Wulff
gu48sah@mytum.de

February 10, 2011

Abstract

This paper outlines current developments and research activities in the field of indoor navigation systems. It describes location techniques and positioning systems, as well as approaches and models to achieve hybrid indoor navigation systems. These hybrid systems are mainly based on a combination of symbolic graph-based and geometric models in order to subdivide the entire navigation task into several subtasks. They should locate and navigate end-users from their particularly present and context-depending point of view in a proverbially optimal way. The dimension of optimality considers several aspects, such as required time and way length. Thus, the optimal way is frequently not simply the shortest one. Therefore, the main purpose of these human-centred approaches and models is to establish a highly intelligent indoor navigation system in accordance with user preferences and given context-aware information within a ubiquitous networking infrastructure. Hence, the claim and goal of such indoor navigation systems are to provide location-based services (LBS). These advanced services, which are rendered by the surrounding environment, facilitate user activities and support people to manage their daily challenges in a smart and unobtrusive way. Furthermore, two existing hybrid indoor navigation systems are introduced and a summary as well as a brief discussion about the past results and an outlook on prospective usage scenarios are given.

1 Introduction and Motivation

During the last few years, the trend to be always online and connected has rapidly increased and taken enormous steps forward in terms of being available [1]. This is on the one hand caused by the demand of individuals, living in a so-called information age, to be further more an active part of a modern society and on the other hand caused and driven by the new technological advances, especially in the area of wireless communication. And, this process is still going on and has definitely not reached one's end so far. For example, within the automotive sector, for years, many automobiles have been equipped with standardised and fixed installed GPS-based outdoor navigation systems and for quite some time, now also portable navigation assistants and systems are available. Since these systems have been well investigated concerning their ability to solve navigation tasks and problems and have also successfully proved themselves in practice, it is only consequent to enhance and transfer them to indoor navigation problems. But it turns out, that indoor navigation problems differ significantly from their outdoor navigation counterparts. This is because of the much more diverse topological structure of indoor environments, so that only few ideas, approaches and knowledge can be extracted and transferred from outdoor to indoor navigation [2]. But taking into account the great advantages and capabilities an indoor navigation system can perform, it is worthwhile to make efforts and enforce the development and design of indoor navigation systems. For example, indoor navigation systems could not only navigate pedestrians through big building complexes, airports, train stations, fairground halls, hospitals, etc., but could also support blind or hampered people in their daily challenge to find an accessible way, and there are more really helpful advantages, which one can consider.

2 Positioning and Navigation Systems

Positioning and navigation tasks basically depend on knowing the user's current location. So, for every navigation system, regardless of whether applied indoor or outdoor, it is important and more or less indispensable to calculate and determine the location with a given precision and within a specified accuracy. This holds particularly, since location is one of the most important components of user context regarding pervasive computing and location technologies have been an important part of ubiquitous computing [3] [4]. Location and position awareness is simply a basic requirement for location-based services (LBS) in context-aware applications on mobile devices [5]. Therefore, in the following Section 2.1 the most common location technologies are described. In Section 2.2 an overview and explanations of various established positioning systems are given and in Section 2.3 two indoor navigation systems are presented.

2.1 Location Technologies and Techniques

In order to accomplish the positioning and location estimation task in known or unknown physical environments, the first step is usually an initial measurement and determination of the distance between a mobile device to be located, and so-called reference devices, whose location is known in advance. In general, distances can be estimated with several methods, for instance, by measuring the time a signal takes between a (moving) object and a reference point, the so-called time-of-flight (TOF) or time-of-arrival (TOA), or by taking into account the attenuation of an emitted signal correlated with increasing distance, and so on. As soon as the distance estimation is accomplished, the basic purpose, namely detecting the position and location of a mobile device in absolute, relative or symbolic form with an accuracy of a priori specified limit and within a user-acceptable convergence time, has to be done. Thereby means the representation in absolute, relative or symbolic form the kind of describing the estimated location, i.e. geographic coordinates are suitable for specifying a precise absolute location which is for example needed for GPS-based cartography, but are not convenient to use in the types of applications that involve reasoning with location information by humans [6, Chap. 7]. For way finding and navigation tasks a description of a location in a relative or symbolic manner is often superior to absolute ones. The relative and symbolic expression "crossing/corner Theresienstraße and Arcisstraße, 50 meters straight ahead and then right at the gate" and especially the "3-D-map" shown in Figure 1 seems to be more familiar and intuitive than the real geographic coordinates, which are by the way 48 09’01.41” degrees north latitude, 11 34’10.51” degrees eastern longitude and 525 meters height.

Figure 1: "3-D-map" / "Picture" from Google StreetView of one of the main buildings of Technische Universität München, as an intuitive example for exploiting the human cognitive ability. Source: http://maps.google.de/maps?hl=en&q=3d%20karte&um=1&ie=UTF-8&sa=N&tab=wl

Besides differences in positioning and location systems with respect to scale and representation, location systems can be further divided in three different classes: 1) in the client-based or self-positioning class, 2) in the network-based or remote positioning class and 3) in the network-assisted class. In the client-based class, the physical location is self-determined by the mobile device of a location system without relying on the network infrastructure, so that only the mobile device itself and not the associated network knows its current location. An example for client-based location systems is GPS. A mobile device equipped with a GPS chip calculates namely its own location using signals received from at least four GPS satellites. In the network-based class, the physical location is remote-determined by the associated network of a location system, i.e. at the server side, using signals continually emitted from the mobile device. Only the network knows the current location of the mobile device and can either use the location for network applications, e.g. mobile tracking, or transmit the estimated location to the mobile device. However, network-based location systems are precarious concerning security and privacy aspects, because the mobile device or more specifically the user does not only know his current location but also what application the network executes and who else knows about his personalised data. An example for such a location system is Active Badge [7] (see Sec. 2.2). Nevertheless both location systems have got their qualifications, which are combined in due consideration with regard to their advantages and disadvantages in the network-assisted class. Therein both the mobile device as well as the associated network compute the current location of the mobile device and accomplish various applications, which in turn complement each other and supply additional information, like e.g. in Assisted GPS [6, p. 289 / 290].

2.1.1 Proximity Sensing

The proximity sensing exploits the closeness of a mobile device to one or more reference points and is thus the simplest location approach as well as technique. The computation of the proximity can be achieved either through a detection of direct physical contact, which is carried out by means of contact sensors, such as pressure, touch and capacitive field detectors, or through a detection of a mobile device being within the coverage of one or more reference points, which is realised by incessantly observing and monitoring the reference points in an assigned network. Although this location technique is easy to implement into communication systems with corresponding reference and/or access points, one great disadvantage of proximity sensing is the dependence on the range of coverage related to the at a time used location technology, like Bluetooth, infrared sensors, Wi-Fi, RFID (radio-frequency identification). Accordingly, one can just say in which region a mobile device is located, but nothing else. Therefore more advanced location techniques and technologies utilize the weighted average of distance information given by more than one reference point locations.

2.1.2 Trilateration

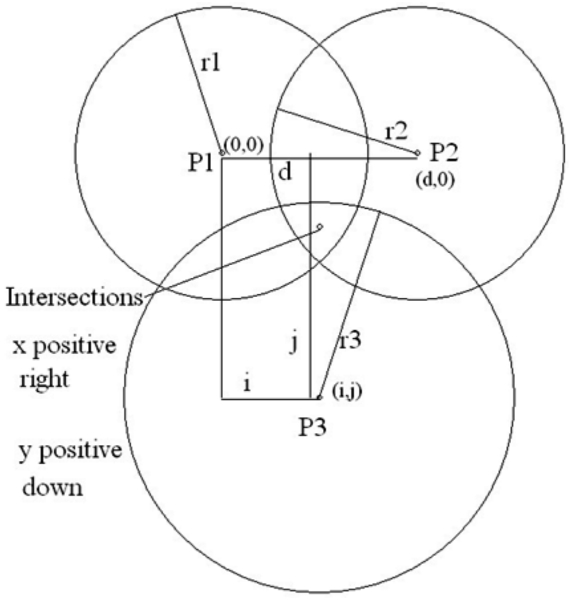
The trilateration technique is a method to calculate the absolute or relative location of a point by measuring the distance between the desired point location or more specifically the object position and some references points, using geometry properties of spheres and triangles (see Fig. 2 (a) and Fig. 2 (b)). In doing so, it is required to measure the distance between the desired point location and a number of reference points, that is one greater than the number of physical space dimensions. For instance, in a 2-D space, it is necessary to have got three non-collinear reference points, and in a 3-D space, like in case of GPS, it is even necessary to get signal information from at least four non-coplanar reference points, whereby the pure measurement of distances is accomplish by several methods, as already mentioned above.

2.1.3 Triangulation

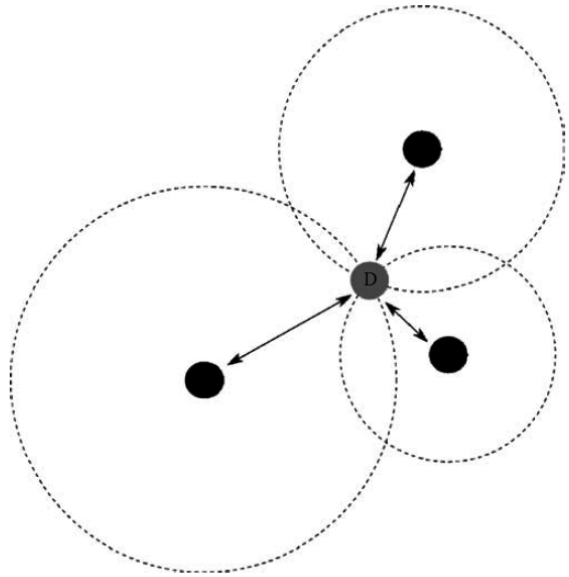
In contrast to the trilateration technique, the triangulation technique calculates the location of a point by measuring angles, instead of distances, to it from known reference points. Exploiting again the geometry properties of spheres and triangles, the desired point location can be assumed and fixed as the third and missing point of a triangle. Hence, in 2-D space, the location of a device can be determined by means of only two, so-called angles of arrival (AOA), whereas the intersection of the two lines, starting from one of the reference points under the specific measured angle of arrival uniquely defines the desired device location (see Fig. 3 (a)). However, for practical applications the triangulation technique relies usually on more than two angles of arrival, and also the distances between the references points are a priori known, in order to reduce the location error significantly [6, Chap. 7].

2.1.4 Hyperbolic Lateration

Another method to calculate the location of a point is the hyperbolic lateration or multilateration technique, which is quite close related to the trilateration technique, presented in Section 2.1.2. Both differ, superficially seen, just on the part of how the distances between a (mobile) device and the corresponding reference points are determined. In fact, the hyperbolic lateration technique estimates the differences between a device and the reference points by measuring the signal arrival times, instead of the signal travel times themselves. Thus, it calculates the desired distances and the point location through the intersection of two or more hyperbolas, which basically arise from the knowledge of the time differences of the signal arrival times (time difference of arrival (TDOA)), which, by contrast, are again quite similar to the triangulation technique or more specifically to the angles of arrival (see Sec. 2.1.3 and Fig. 3 (b)). Consequently, it becomes obvious that several location techniques, especially mixed forms, can be constructed. However, in this paper only two more popular location techniques are presented in the following (see Sec. 2.1.5 and Sec. 2.1.6).

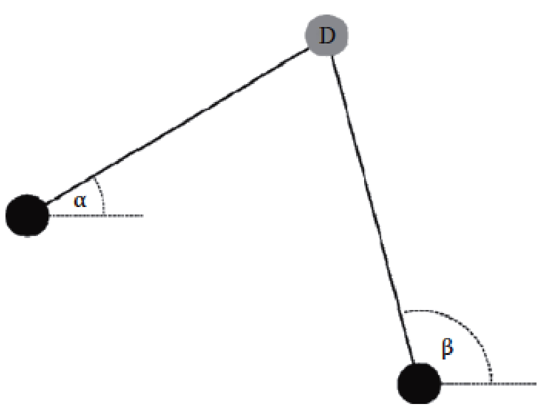


(a) Approach of trilateration in 2-D; showing the 3 sphere centres, P1, P2 and P3; their x,y coordinates (z=0); and the 3 sphere radii, r1, r2 and r3. The two intersections of the three sphere surfaces are directly in front and directly behind the point designated intersections in the z=0 plane. Source: http://en.wikipedia.org/wiki/Trilateration

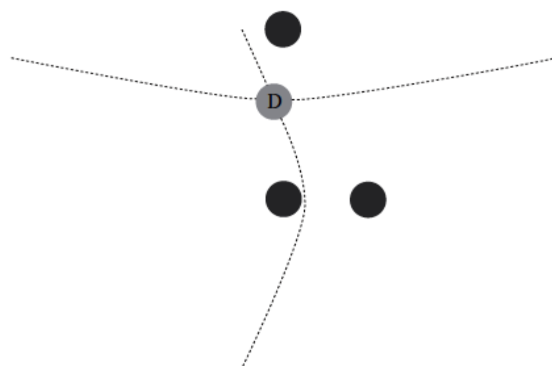


(b) Trilateration in 2-D. The black dots represent reference points. The gray dot represents the location of the device [6, p.292].

Figure 2: Positioning by Trilateration.



(a) Triangulation in 2-D. The black dots represent reference points. The gray dot represents the location of the device [6, p.295].



(b) Hyperbolic lateration in 2-D. The black dots represent reference points. The gray dot D represents the location of the device [6, p.294].

Figure 3: Positioning by Triangulation and Hyperbolic Lateration.

2.1.5 Fingerprinting

Location Fingerprinting is a location approach that relies on pattern matching algorithms in terms of comparing the received signal strengths at any given location over time (temporal stability) as well as at any specified location (spatial variability). For this, a training phase, in which measurements are taken in an arbitrary close grid pattern to establish a database, is inevitably necessary, before the localisation algorithm is able to start. The location approach itself is then, for example, based on a matrix correlation algorithm or diverse other matching algorithms, which are used to search the database for the best match. Typically, the result (best match) of the matching algorithm is very accurate and precise, but takes, beside the additionally required training phase, a relatively long time to implement and perform.

2.1.6 Dead Reckoning

Dead reckoning is a location technology that estimates the location of a mobile device in relation to an initially known location, i.e. the current location is calculated through the previous location and from the estimated or even interpolated velocity, travel direction and elapsed time. As a consequence, this location approach implicates the disadvantage that the location error probability increases exponentially over time. Thus, dead reckoning is normally just applied in modern inertial navigation systems or either to refine the location estimates of another location systems or to support other location systems, when they are temporary unavailable [6].

2.2 Positioning Systems

Once the location technologies and techniques are conceived and established, the next step to develop a real working positioning system is a general challenge to trade off the positive and favoured properties of location technologies and techniques against the negative ones, because there is no perfect positioning system and only an approximation can be achieved. Accordingly, every concept of building and installation positioning systems should be in principle based on intended applications and should consider particular scenarios, like outdoor and indoor environments. In order to evaluate the pros and cons of different positioning systems, the annexed Table 1 provides an overview of common positioning systems and highlights several characteristics, such as the location type, the resolution / accuracy, the infrastructure requirements, the location data storage, the spectral requirements and the location system type, which distinguish different positioning system solutions.

2.3 Indoor Navigation Systems

For indoor navigation tasks, at present, exist a plenty of commercially available systems, like Ubisense, an ultra-wide-band-based high resolution system for indoor and outdoor environments (see Tab. 1), or MagicMap, a WiFi-based indoor navigation system [15]. Furthermore several other indoor navigation systems, solely based either on ultra-wide-band (UWB), WiFi/WLAN, Bluetooth or infrared technology, exist. Recently badge-based systems, which combine two or more of these technologies are increasingly available and used, such as Topaz, that combines Bluetooth and infrared-based technologies (see Tab. 1) and Elpas, that combines infrared, low frequency and radio frequency technologies [16]. Other indoor navigation systems are some two-dimensional or more precisely matrix barcode systems, grounded on the already standardised Quick-Response (QR)-Codes (ISO/IEC 18004), which have grown up and established, and up to now retain on the market. Beside these indoor positioning systems presented above and listed in Table 1, there also exist further potentially positioning and tracking systems, like radio-frequency identification (RFID) and Near Field Communication (NFC) as well as vision-based tracking systems, such as EasyLiving [17] and Vicon [18], a lately common camera-based motion tracking system. Although many potentially indoor navigation systems are on hand, concepts, buildings and installations of deployed hybrid indoor navigation systems are quite rare and still in their infancies, yet, being a dynamic part of current research activities. Two of these more or less research-driven hybrid indoor navigation systems are introduced in Section 4.1 and Section 4.2 respectively.

3 Hybrid Models for Indoor Navigation

Indoor navigation systems, which incorporate user preferences and given context-aware information, require hybrid models to calculate possible paths between two given locations efficiently and to provide further advanced and smart services, as desired. Manifold variation of conditions people may encounter in indoor environments causes that hybrid models for indoor navigation have not only to solve a modest way finding task but also take into account situation-related circumstances and conditions as well as particular wishes of the users. This considerably complicates the way finding task, because the optimal path or more precisely the optimal way finding algorithm is anyway influenced by various factors and dimensions of optimality. To overcome these challenges and obstacles first and foremost hierarchically structured graph models are conceived and build up, which are described in Section 3.1. Moreover in Section 3.2 pervasive adaptations, such as physical capabilities and limitations, are shown. In Section 3.3 some specific path calculation and selection algorithms are explained and in Section 3.4 graphical presentation schemata are introduced.

	Location Type	Resolution, Accuracy	Infrastructure Requirements	Location Data Storage	Spectral Requirements	Location System Type
Active Badge	Symbolic Indoor	Room level	IR Sensors and customs tag	Central	IR	Custom active tagging
ActiveBat, Dolphin	Absolute Indoor	3 cm, 90%	Ultrasonic (US) receivers and transmitters	Central	30 kHz ultrasound and 900 MHz RF	Custom active tagging
ActiveFloor	Symbolic Indoor	1 m, 91%	Custom floor tiles	Central	Load sensor	Passive
Airbus	Symbolic Indoor	Room level, 88%	Single sensor in HVAC	Central	Pressure sensor	Passive
Beep	Absolute Indoor	0.4 m, 90%	Acoustic sensors and transmitters	Central	4.01 kHz	Custom active Tagging
Cricket	Absolute Indoor	3 cm, 90%	US receivers and transmitters	Local	30 kHz ultrasound and 900 MHz RF	Custom active tagging
DECT	Absolute Indoor/Outdoor	5 m, 55% 15 m, 90%	Base stations and mobile devices	Central	1.8 GHz (1.9 GHz)	Active Tagging
Ekahau	Symbolic Indoor	1 - 2m, 90%	3 (or more) WiFi APs	Central	2.4 GHz RF	Active tagging
GPS	Absolute Outdoor	10 m, 50%	GPS receiver	Local	1500 MHz RF	Custom active tagging
HORUS	Symbolic Indoor	1.4 m, 90%	WiFi APs	Local	2.4 GHz RF	Active tagging
Landmarc	Absolute Indoor	1 - 2 m, 50%	RFID reference and tracking tags and readers	Central	308 MHz	Custom active tagging
MotionStar	Absolute Indoor	1 cm, n/a	Custom sensors and transmitters	Central	Pulsed DC magnetic fields	Active Tagging
PlaceLab (GSM)	Symbolic Indoor/Outdoor	20 m, 90% 5 m, 50%	Existing GSM towers	Local	900 - 2000 MHz RF	Active tagging
PlaceLab (WiFi)	Symbolic Indoor/Outdoor	20 m, 50%	Existing WiFi APs	Local	2.4 GHz RF	Active tagging
PowerLine Positioning (PLP)	Symbolic Indoor	2 m, 93% 0.75 m, 50%	2 plug-in module and custom tag	Local or central	300 - 1600 kHz RF	Custom active tagging
RADAR	Symbolic Indoor	6 m, 90% 2 - 3 m, 50%	3 - 5 WiFi APs	Local	2.4 GHz RF	Active tagging
SmartLocus	Absolute Indoor	2 - 15 cm, 50%	US receivers and transmitters	Local	40 kHz ultrasound and 900 MHz RF	Custom active tagging
SpotOn	Absolute Indoor	1 - 3 m, n/a	RFID tags and readers	Central	916.5 MHz	Custom active tagging
Topaz	Symbolic Indoor	2 - 3 m (Room level)	Bluetooth and IR tags, APs and servers	Central	Bluetooth (2.4 GHz) and IR	Active Tagging
Ubisense	Absolute Indoor/Outdoor	15 cm, 90%	Custom sensors and tags	Central	2.5 GHz and 6 - 8 GHz wideband RF	Custom active tagging
Vision	Absolute Indoor/Outdoor	1 m, 50 - 80% (varies by camera density)	Multiple cameras	Central	RF for wireless cameras	Passive
Wherenet	Absolute Indoor/Outdoor	2 - 3 m, 50%	Location antennas and tags	Central	2.4 GHz (UHF)	Active Tagging

(Note that the accuracy is reported as a percentile.)

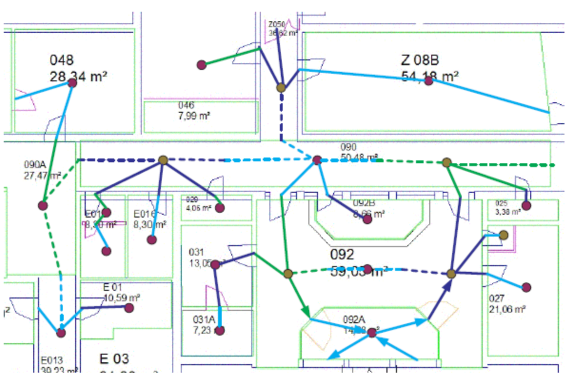
Table 1: Location and Positioning Systems (according to [6, p. 301], [8], [9], [10], [11], [12], [13] and [14])

3.1 Graph Model for Indoor Environments

In order to fulfil and meet the requirements of optimal way finding tasks, it is necessary or at least helpful to divide the way finding and presentation task into several steps. The first step is usually to calculate possible paths between the current location and the final destination, whose attributes are afterwards, in the next steps, valuated and weighted by many diverse influence factors, such as distance, crowdedness and luminosity. Because it turns out, that the most efficient way finding solutions use shortest-path algorithms in graphs, and graphs also provide relatively intelligible possibilities to integrate weighted influence factors, graph models are commonly used to solve indoor navigation tasks in pervasive systems. These graph models are basically hierarchically structured and enriched with extra qualitative and quantitative information assorted to model two- or even 2.5-dimensional structures of floor plans and multiple storeys. Accordingly, these graph models partition two-dimensional indoor environments into cells and represent them as labelled graph nodes and edges in different levels of abstraction [2]. Not until then, based on the developed hierarchical graph models, for example, route description and graphical way presentation problems are approached in the next steps. Thus, graph models for indoor environments are the initial and central point to establish user-centred navigation systems.

3.1.1 Decomposition

In order to develop a hierarchical graph model for indoor environments a partition and decomposition of two-dimensional areas into cells has to be done. These cells, such as rooms and corridors, are then represented as nodes in the graph model, whereas doors and other pass ways between rooms and corridors are modelled as edges. Thus, a direct one-to-one mapping from a floor blueprint to a graph model structure can be achieved. Yet, this holds only for simple and small rooms and corridors, because larger and more diverse rooms and corridors corresponding to larger buildings, such as airports or university main halls, cause difficulties in partitioning and decomposing given areas. Therefore, such large rooms and corridors are split and divided into several, non-overlapping and disjoint cells, which are connected to their adjacent cells just by a link. Figure 4(a) shows such a partition and decomposition of a blueprint of a university building into cells by overlaying the floor plan with a suitable graph structure. The rooms and corridors are thereby not even split into several cells due to their length and size, but also due to their concavity and functionality.



(a) With Cell Centres and a Path System Overlaid Floor Plan that accentuates the diversity and complexity of decomposing larger rooms and corridors into non-overlapping and disjoint cells. [2].

Granularity \ Model Element	Graph	Vertex	Edge
Level 4 (coarsest)	City	Building	StreetNetwork
Level 3	Building	Storey, *Staircase*, *Elevator*	
Level 2	Storey	Wing, Room, Corridor	Door, Window, Ladder, Ramp, Stairs
Level 1	Wing	Room, Corridor	
Level 0 (finest)	Room, Corridor	PartOfRoom, PartOfCorridor	

(b) Relations between Hierarchy Level and Graph Elements [2]

Figure 4: Hierarchical Graph Model for Indoor Environments

3.1.2 Hierarchical Graphs

Behind every direction and path explanation how to get to a particular location and with that, also behind every indoor navigation task, stands a multi-level hierarchical model of the environment. Hence, it is only consequent to transfer these two- or multi-level models into hierarchical graph models with different abstraction levels, which support and provide a hierarchical planning of indoor navigation tasks. Such a graph model with its levels and identifiers within the hierarchy is illustrated in Figure 5. These levels are needed to model differences in the granularity between entire buildings, storeys, wings, corridors and rooms, whereby the identifiers link the graphs at different levels together and thus, enable to access higher levels from lower levels and vice-versa. This linking is basically realised by different representations or more precisely generalisations and refinements of same areas at different graph levels. For example, a storey of a building can be represented as an entire graph at a certain level n, whereas the same storey and the appropriate entire graph can be indeed represented just as a node in a graph at a higher level n + 1. In addition, edges and so-called "interface nodes" assist the connections and transitions between different graph model levels. Thereby upcoming relations between hierarchy levels and graph elements are shown in Figure 4(b). Caused by this, way finding algorithms by means of shortest path algorithms, which require just such a graph model and a cost function to measure the geometrical distance between two places, can be implemented relatively easily. Nonetheless, in order to generate and provide smarter way finding algorithms with regard to user preferences and context-aware information, more sophisticated cost functions and path descriptions are necessary. For this purpose, the graphs must be enriched with a lot more semantic information. This additionally class-dependent information can be simply attached to the nodes and edges within the graph model. Thus, indoor environments are finally well classified and projected onto hierarchical graph models [2].

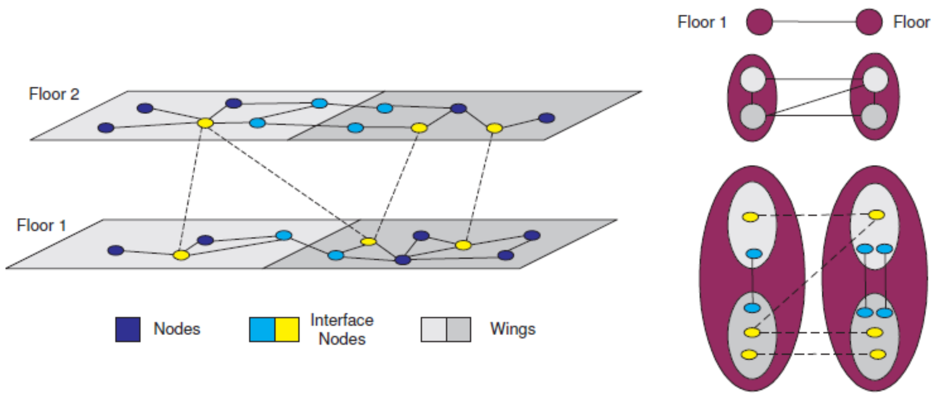


Figure 5: Hierarchical Graph with its levels and identifiers at different abstraction levels to support and provide a hierarchical planning of indoor navigation tasks [2].

3.2 Model for a user-centric and pervasive Adaptation

In addition to hierarchical graph models for indoor navigation tasks, further models for a user-centric and pervasive adaptation of the aforementioned way finding algorithms must be considered and integrated to realise a real user-centric navigation paradigm for indoor environments. These additional models represent hereby a kind of collection of specific user capabilities and limitations, like in personalised user profiles. In order to classify user capabilities and subsequently to create user profiles with classified and weighted attributes, surveys based on physical, cognitive and perceptual capabilities and limitations, user preferences and location access rights have to be carried out [1]. The physical, cognitive and perceptual capabilities are measured with the help of classification systems, such as the international classification system and framework respectively for measuring health and disability (ICF), developed by the World Health Organization (WHO) [16]. The user preferences, however, are really hard to measure and to quantify regarding just contextual information, because e.g. the beauty of a path involves several aspects, such as colour combination, layout, among others. That is why, user preferences usually just rely on direct user inputs, and also personal location access rights can be seen as direct inputs into the user profile system. Hence, all the additional information about user characteristics is resided in user-centric models and stored in user profiles, which can be individually and dynamically updated by the user to react on different situations and unforeseen events.

3.3 Path Calculation and Selection

In order to calculate and select the best adapted path regarding user preferences, rating and evaluation methods are used, which describe feasible and walk able paths by individual preference attributes. These attributes are in turn used to measure path preference strengths through value functions and thus, to calculate an overall path evaluation value by aggregating the resulting single attribute evaluation values for every path. The path with the highest overall evaluation value is finally chosen as the most suitable one for a specific user, with the path length being just one assessed and weighted attribute of many others. For instance, such a rating and evaluation method for solving the abovementioned optimal path calculation and selection problem is introduced and presented in [16]. The therein described method, the Assessment of Preferences with Simple Multiattribute Rating Technique (SMART), can be formalised as follows:

> „Let P be the set of all feasible paths, then P = $\{p|p$ is a path over the location model$\}$ and a_n all attributes with importance weights ω_n and value functions υ_n. Then, the optimal path $p \in P$ maximizes $eval(p) = aggregate(\omega_i(\upsilon_i(a_i(p))))$ for i = 1,. . . , number of attributes "

and defined as well as conducted within the following five steps:

1. Define feasible paths and attributes for evaluation.
2. Assign the importance weights to the attributes.
3. Assess the path attributes separately for every feasible path.
4. Calculate evaluation value by aggregating all attribute assessments under consideration of importance weights.
5. Select the path with the best evaluation value.

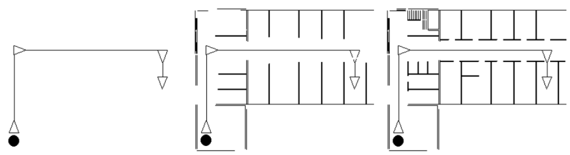
Beside this particular method and technique respectively, many other methodical strategies and analytic procedures can be considered, and already exist. Those often separate and divide the optimal path calculation and selection problem just in a three- or even two-phase problem, like in [1].

3.4 Graphical Presentation

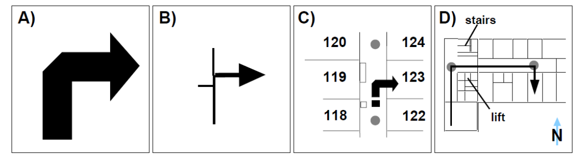
As soon as the optimal path for an indoor navigation task is calculated and selected, the question, how to present the obtained results with human-understandable and intuitive path and route descriptions, arises. This can only be answered by taking into account that the form of route descriptions and presentations must be adapted to the available resources of the environment and current situation a user is in. Hence, in order to produce a coherent and cohesive route presentation, it is crucial to observe and regard technical resources, such as speed, bandwidth and screen resolution limitations of the mobile device, as well as cognitive resources and restrictions of the user [19]. But not only technical and cognitive resources affect the form of route description but also the precision and accuracy of the user position and orientation has a significant impact. As a consequence of these various dependences, graphical presentation schemata and different presentation techniques, such as allocentric 2-D-maps, egocentric walkthroughs (2-D-sketches) and 3-D-visualisations, among others, to handle these resources and consequently to provide the most suitable route presentation are developed. The main aspects of such a graphical presentation schema and presentation technique, introduced in [20], are described below.

3.4.1 Graphical Schemata

Technical and cognitive resources and restrictions as well as differences with respect to the precision and accuracy of user position and orientation information require different kinds and forms of graphical presentations for way descriptions. While technical restrictions mainly influence the decision what kind of way descriptions are technically feasible to present on mobile devices, cognitive restrictions and different qualities of position and orientation information influence mainly the decision how way descriptions should be presented. Thus, caused by technical limitations the transmission of already generated way descriptions is often divided into several steps, as shown in Figure 6(a). Therein, the way description becomes more and more detailed in every step. Beside such an incremental transmission of vector graphics, also parallel or sequential transmissions are used, if they are technically feasible. Furthermore, somewhat apart from technical restrictions, not only time-dependent transmissions, including the aforementioned concepts of parallel, sequential and incremental transmissions, but also logical transmission interrelationships, including concepts of alternative, conditional and additional transmissions, are developed and commonly used [20]. However, in order to present way descriptions well-adapted with respect to cognitive restrictions and different qualities of position and orientation information, graphical presentations for way descriptions change in the effect that they become coarser, if the quality and knowledge about the user's current position and orientation decreases. Hence, if the user position and



(a) Incremental transmission of vector graphics: First an arrow depicting the path is shown, then the floor plan is transmitted in several steps, longest lines first [20].



(b) Four different graphical way description schemata that depend on the quality of orientation and position information [20].

Figure 6: Graphical presentations for way descriptions.

orientation is well known, the graphical way description can be presented very exactly and precisely, as e.g. with a simple arrow. But, if the quality of the user position and orientation information declines, the graphical way description must be modified and enhanced with additional information, such as landmarks, to provide any longer enough information to navigate the user. An example for such a modification and enhancement of graphical way description schemata is illustrated in Figure 6(b), where the knowledge about the current user position and orientation decreases from subpart/subfigure A) to D) and consequently, the amount of presented information increases from A) to D).

3.4.2 Presentation Techniques

Besides the abovementioned differences of graphical presentation schemata depending on the quality of position and orientation information, there also exist great differences of presentation techniques to display a way description on a mobile device in a suitable manner. Yet, both parts are closely related together, because the selected graphical presentation schema predetermine and predesignate the appropriate presentation technique significantly. For instance, a well known user position and orientation leads usually to a precise and simple graphical way description, such as the in Figure 6(b) A) depicted arrow. This leads in turn to certain presentation techniques by excluding bird's eye views, allocentric maps and 3-D-visualisations, among others. A full and detailed description of different presentation techniques and their ability to present different kinds of knowledge can be found in [19] and [20].

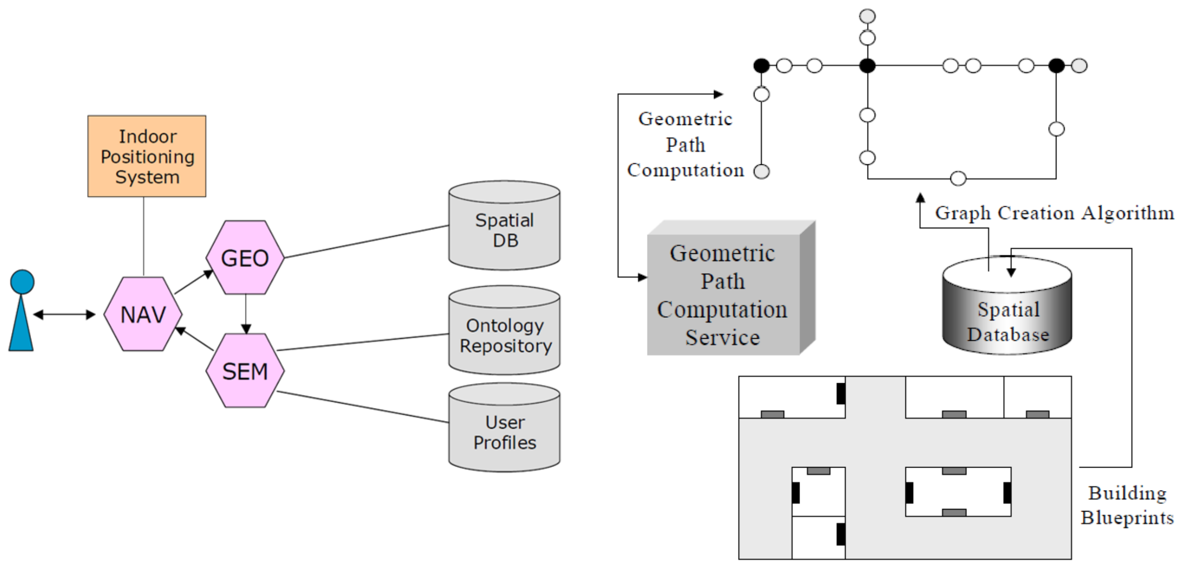
4 Hybrid Indoor Navigation Systems

Although many positioning systems for outdoor as well as indoor environments are commercially available, as described in Section 2.2 and Section 2.3, deployed indoor navigation systems with a consistent concept and structure are quite rare. Especially, those indoor navigation systems, which are purely user-centred and based on a hybrid model that combines symbolic graph-based or semantic models with geometric ones, are very seldom and still need to be improved. But nonetheless in Section 4.1 and Section 4.2 already developed hybrid indoor navigation systems are delineated.

4.1 Semantic Indoor Navigation System OntoNav

One example for a hybrid indoor navigation system is the Semantic Indoor Navigation System OntoNav, an integrated and purely user-centred navigation system for indoor environments. The OntoNav system relies mainly on an indoor navigation ontology (INO), which supports both the path searching and the presentation tasks of a navigation system. Hence, it provides several location-based services (LBS) beyond common navigation tasks and attempts to combine single problem solution services in a smart and integrated way. For that reason, the OntoNav system consists of the following building blocks:

- Navigation Service (NAV): is the main interface between the user and the system.
- Geometric Path Computation Service (GEO): computes all geometrical paths by means of a graph-traversal searching algorithm.
- Semantic Path Selection Service (SEM): filters the in GEO already calculated navigation paths and selects the best traversable one with respect to user capabilities and preferences.



(a) Overview of the OntoNav architecture that illustrates the composition of the main building blocks and their interrelationships [1].

(b) The GEO Service functionality to compute all geometrical paths by means of a graph-traversal searching algorithm [1].

Figure 7: The Semantic Indoor Navigation System OntoNav.

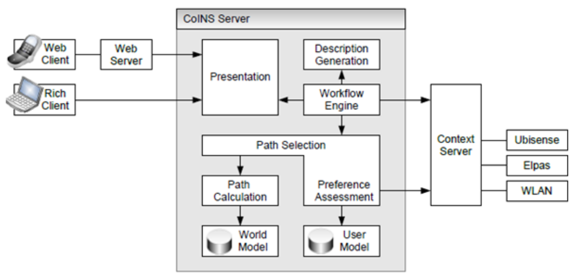
An overview of the OntoNav system architecture is given in Figure 7(a), from which the composition of the main building blocks and elements as well as their interactions between each other can be seen. In Figure 7(b) the functionality of one of the three main service building blocks, the Geometric Path Computation Service (GEO), is pictured. Beyond these three main service building blocks, the OntoNav system also provides an optimisation service element and moreover even an additional navigation-aiding module (NAM), which supports users to detect potential deviations and reschedule the initially planned path [1].

4.2 Context-Aware Indoor Navigation System (CoINS)

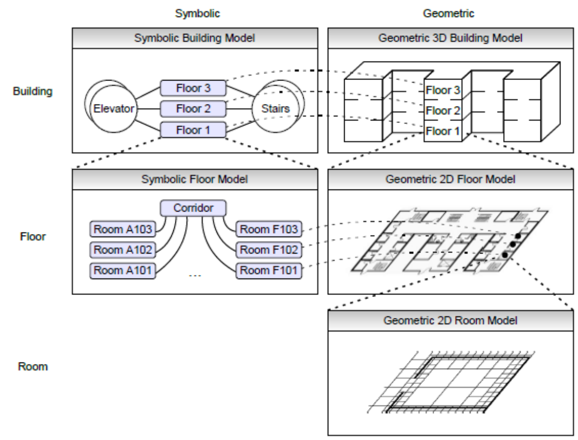
Another already existing hybrid indoor navigation system is the Context-Aware Indoor Navigation System (CoINS), an integrated and user-centred navigation system for indoor environments, just like the abovementioned OntoNav system. Also identically in both navigation systems is the hybrid model approach, i.e. a combination of symbolic graph-based and geometric models, to subdivide the entire navigation task into several subtasks, such as in Section 3.1 already described. For this purpose, CoINS consists of several various building blocks and components within a multilayered system architecture, which is illustrated in Figure 8(a). Thereby the main components are:

- the ’Presentation’ component to depict and display intuitive route descriptions
- the ’Path Calculation’ component to calculate and determine all possible routes
- the ’Path Selection’ component to exclude possible routes by use of the Preference Assessment and subsequently to select the most suitable route
- the ’Workflow Engine’ to control the whole navigation process

Moreover for an efficient navigation process it is necessary to dynamically assign these system components to their currently appropriate CoINS model. For instance, the Path Calculation component, which considers only the geometric aspects of the navigation task, requires self-evidently the geometric model. But, it also requires the symbolic graph-based model to create a first, coarse plan of the route by using a graph searching algorithm. Hence, a transformation from the symbolic to the geometric model or more precisely from symbolic to geometric location coordinates and vice-versa must be possible. Thus, the overall hybrid world model of CoINS, which is illustrated in Figure 8(b), enables an unimpeded transition between the symbolic graph-based model and the geometric model to provide human cognitive ability adapted route descriptions.



(a) Overview of the CoINS architecture that illustrates the main and most important elements and their interactions among each other [16].



(b) CoINS World Model that depicts the symbolic as well as the geometric model and elucidates the transition between both models [16].

Figure 8: The Context-Aware Indoor Navigation System (CoINS).

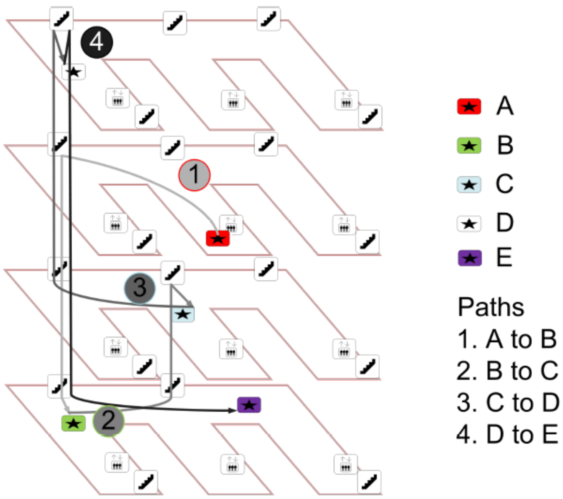
5 Results and Discussion

Human way finding and navigation as well as the associated conveyance of way descriptions to guide and orient others has been studied from many different perspectives and recorded in several navigation models and algorithms, which are generally accepted and used in common outdoor navigation systems [16]. Although outdoor navigation systems are practically standard in daily life, indoor navigation systems are still on aforementioned grounds of conceptual and technical difficulties quite rare. As a result of this, up to now only a few measurements and evaluations of deployed indoor navigation systems are carried out and consequently, only a few well-founded and solid evaluation results are currently available. One example for a already valued and assessed indoor navigation system is the in 4.2 introduced and presented Context-Aware Indoor Navigation System (CoINS). A review of the main and most important results of this evaluation can be found in Section 5.1 and more detailed in [16]. Section 5.2 reveals future prospects for indoor navigation systems as well as further potential applications, and finally in Section 5.3 a summary is given.

5.1 Evaluation of existing systems

Measurement and evaluation results of existing indoor navigation systems are rare. Therefore, the following remarks on conducted experiments refer to the Context-Aware Indoor Navigation System (CoINS). In order to prove and examine the functionality of CoINS its developers have carried out two different experiments for a first evaluation on principles. The first experiment was applied to gather information about contextual user preferences and to find out criteria that considerably influence the selection of a particular route. Therefore, 8 subjects were encouraged to walk from a given origin to particular destinations within a familiar building, while they were additionally asked to fulfil simple tasks that implied walking between the given destinations. Meanwhile the actually selected and followed routes as well as the possible ones were recorded and stored in separate maps and then analysed in combination with an additional questionnaire on the criteria the subjects used in selecting a route. For the second experiment exactly the same destinations and simple tasks as in the first experiment were used, but 9 other subjects were chosen. Also, the gathered user preference profile informations were used for the CoINS Path Calculation component and algorithm respectively. Finally, the paths suggested by CoINS were compared to the ones actually selected and followed by the subjects. The results of this comparison can be seen in Figure 9, which show that 86,1% of the paths followed by the subjects match and correspond with those calculated and suggested by CoINS. Although these results are encouraging, nonetheless this evaluation must be extended to larger indoor environments and ought to be just seen as a starting point for further evaluation series in order to achieve profound verifications of indoor navigation systems in general.

(a)



(b)

Subj.	M/F	A to B Path Chosen	B to C Path Chosen	C to D Path Chosen	D to E Path Chosen
1	M	Y	Y	Y	Y
2	M	Y	Y	Y	Y
3	M	Y	Y	N	Y
4	F	N	Y	Y	Y
5	F	Y	Y	Y	Y
6	F	Y	Y	N	Y
7	F	Y	Y	Y	Y
8	M	N	Y	N	Y
9	F	Y	Y	Y	Y

Figure 9: (a) Shows the structure of the Piloty building at the University of Darmstadt, and the paths proposed for evaluation. (b) Table shows for each participant and path, whether the path followed by the subject matches the path proposed by the Context-Aware Indoor Navigation System (CoINS). [16]

5.2 Outlook - Example and Vision

„Sue is rushing through the door of the huge office complex, into the empty lobby, . . . "Hello Mrs. Walker" says a voice from the building, while a small puddle is forming on the marble floor around her feet. "Can I take you to your 1PM appointment with Mr. Grey? You are well in time, so if there's anything else I can do for you..." "Coffee" she nods. "That'd be a bliss..." A projection of a boy scout appears in front of her, smiles, and gestures to follow her. "So let me take you to our coffee machine first." . . . A paper cup in her hand she follows the little guy to the elevator. "I'll send you to the 63rd floor. You'll just have to follow the arrows on your bicorder then" are his last words while the elevator is closing with a soft hiss. Upon entering the nested system of hallways, Sue takes out her bicorder and sees a red arrow appear on its display. She follows its direction which is changing every few corners. Just before she reaches Mr. Grey's office, she drops the empty cup into the tray marked on a small map on the bicorder screen. "You're well in time, Mrs. Walker" comes the voice from behind the desk. "Glad you found it!" . . . " [20]

Scenarios like this are still up in the air and sounds commonly like dreams of the future, which are maybe possible in 15 or 20 years, but nonetheless also nowadays pervasive and ubiquitous computing systems are available and become more and more part of our daily life. Indoor navigation systems are already used in a few particular applications, and development and research activities regarding indoor navigation systems, which have been increased for several years, grow further. It can be expected that in the near future indoor navigation systems will be commercially available with a wide distribution and reach technically and economically a standard comparable with existing outdoor navigation systems.

5.3 Summary

Beside some general location techniques and existing positioning systems, in this paper user-centred hybrid models for indoor navigation systems have been described. Especially, hierarchical graph models, path calculation and selection algorithms, and graphical presentation techniques have been introduced. Furthermore two existing hybrid indoor navigation systems, the Semantic Indoor Navigation System OntoNav and the Context-Aware Indoor Navigation System (CoINS), have been presented and compared with respect to their architecture and functionality. In doing so, it turns out that both systems are integrated and purely user-centred. Both systems also use a hybrid model approach, which means a combination of symbolic graph-based and geometric models, to subdivide the entire navigation task into several subtasks, such as path calculation, selection and presentation. In order to prove the correctness and verify the integrity and sufficiency of these hybrid systems and model approaches respectively, a first evaluation of CoINS has been carried out. The obtained results show that indeed a huge number of paths actually followed by the users or more precisely by the subjects match and correspond with those calculated and suggested by CoINS. Accordingly, the main objectives of user-centred hybrid indoor navigation systems are almost achieved, what gives rise to expect that indoor navigation becomes an incipient and emergent technology.

References

[1] Tsetsos, V., Anagnostopoulos, C., Kikiras, P., Hasiotis, P., Hadjiefthymiades, S.: A human-centered semantic navigation system for indoor environments. (July 2005) 146 – 155 1, 3.2, 3.3, 7(a), 7(b), 4.1

[2] Lorenz, B., Ohlbach, H., Stoffel, E.: A hybrid spatial model for representing indoor environments. Web and Wireless Geographical Information Systems (2006) 102–112 1, 3.1, 4(a), 4(b), 3.1.2, 5

[3] LaMarca, A., de Lara, E.: Location systems: An introduction to the technology behind location awareness. Synthesis Lectures on Mobile and Pervasive Computing **3**(1) (2008) 1–122 2

[4] Norman, A., Schilit, N., Want, R.: Context-aware computing applications. In: Proceedings of the Workshop on Mobile Computing Systems and Applications, 1999. (1999) 2

[5] Genco, A., Sorce, S.: Pervasive Systems and Ubiquitous Computing. 1st edn. WIT Press (2010) 2

[6] Krumm, J.: Ubiquitous Computing Fundamentals. Chapman & Hall/CRC (2009) 2.1, 2.1, 2.1.3, 2(b), 3(a), 3(b), 2.1.6, 1

[7] Want, R., Hopper, A., Falcao, V., Gibbons, J.: The active badge location system. ACM Transactions on Information Systems (TOIS) **10**(1) (1992) 91–102 2.1

[8] Liu, H., Darabi, H., Banerjee, P., Liu, J.: Survey of wireless indoor positioning techniques and systems. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews **37**(6) (2007) 1067–1080 1

[9] Gu, Y., Lo, A., Niemegeers, I.: A survey of indoor positioning systems for wireless personal networks. Communications Surveys & Tutorials, IEEE **11**(1) (2009) 13–32 1

[10] Mandal, A., Lopes, C., Givargis, T., Haghighat, A., Jurdak, R., Baldi, P.: Beep: 3D indoor positioning using audible sound. In: Consumer Communications and Networking Conference, 2005. CCNC. 2005 Second IEEE, IEEE (2005) 348–353 1

[11] Kranz, M., Fischer, C., Schmidt, A.: A comparative study of DECT and WLAN signals for indoor localization. In: 8th Annual IEEE Int'l Conference on Pervasive Computing and Communications (PerCom 2010), Mannheim, Germany, IEEE (Apr 2010) 235 –243 1

[12] Youssef, M., Agrawala, A.: The Horus location determination system. Wireless Networks **14**(3) (2008) 357–374 1

[13] Ni, L., Liu, Y., Lau, Y., Patil, A.: LANDMARC: indoor location sensing using active RFID. Wireless Networks **10**(6) (2004) 701–710 1

[14] Hightower, J., Want, R., Borriello, G.: 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 (2000) 1

[15] Ibach, P., Stantchev, V., Lederer, F., Weiß, A., Herbst, T., Kunze, T.: WLAN-based asset tracking for warehouse management. In: Proceedings of the IADIS International Conference e-Commerce, Citeseer (2005) 2.3

[16] Lyardet, F., Szeto, D., Aitenbichler, E.: Context-aware indoor navigation. In Aarts, E., Crowley, J., de Ruyter, B., Gerhäuser, H., Pflaum, A., Schmidt, J., Wichert, R., eds.: Ambient Intelligence. Volume 5355 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2008) 290–307 10.1007/978-3-540-89617-$3_1 9. 2.3, 3.2, 3.3, 8(a), 8(b), 5, 9$

[17] Krumm, J., Harris, S., Meyers, B., Brumitt, B., Hale, M., Shafer, S.: Multi-camera multi-person tracking for easyliving. vs (2000) 3 2.3

[18] Vicon: Vicon motion systems. http://www.vicon.com (January 2011) 2.3

[19] Kray, C., Elting, C., Laakso, K., Coors, V.: Presenting route instructions on mobile devices. In: Proceedings of the 8th international conference on Intelligent user interfaces, ACM (2003) 117–124 3.4, 3.4.2

[20] Butz, A., Baus, J., Krüger, A., Lohse, M.: A hybrid indoor navigation system. In: IUI '01: Proceedings of the 6th international conference on Intelligent user interfaces, New York, NY, USA, ACM (2001) 25–32 3.4, 3.4.1, 6(a), 6(b), 3.4.2, 5.2

Indoor Navigation 2

Mathieu Cattenoz
mathieu.cattenoz@mytum.de

February 10, 2011

Abstract

This paper is a brief overview of indoor navigation relying on related studies in order to give to the reader a global knowledge about this subject. We describe how to acquire a virtual representation of the environment by software modelling and also with a method of Simultaneous Localisation and Mapping (SLAM) to map on-site with a camera. A graph is then generated from the original representation. Different techniques for locating the user are described: visual identification, proximity detection, mapping of received signal strength, measurement of ranges or angles and electromechanical integration. We further report how to compute the optimal path depending on several parameters and how to communicate routing instructions. Different perspectives and display methods are proposed. Finally we deal with technical considerations on the user terminal.

1 Introduction

Who did never get lost in a building when looking for a room for an important appointment? Imagine now that your phone displays on its screen your current position in a 3-dimension view of the room and tells with vocal instructions in real time, when and where you have to turn in the corridors. In such situation, you would have arrived on time and serenely at your appointment. With the spreading of smartphones and the rising demand on interactive services, indoor navigation has never been so actual and accessible. The range of applications is wide and covers important social issues like giving more autonomy to blind persons in their movements and helping people finding a safe and fast evacuation route in case of an emergency, but also entertainment and commercial purposes like guided tours in a museum and guidance of customers in a shop.

For road navigation the user follows roads seen as lines in a plane and can easily localised his position with GPS-satellite signals. It is more complex with indoor navigation especially because the reception of satellite signal is difficult and the navigation environment comprises entire rooms on defferent floors, it is therefore necessary to proceed differently. Many techniques for indoor navigation come from other scientific domains like robotics and artificial intelligence, in particular the techniques to localise someone, map an environment and compute a path in a graph, and like multimedia techniques for the display of indoor environment on screen.

We present in this paper the different steps involved in indoor navigation. The information is mainly taken from the existing scientific literature. Section 2 presents the methods to acquire and represent the environment and then section 3 deals with the technologies which allow to get someone's position in a building. The methods to compute and display the path are described in section 4 and finally section 5 deals with the technical characteristics of the user navigation device.

2 Representation of the Environment

2.1 Acquisition of the Environment

The system has to localise the position of the user in the known environment. The first step is thus to previously acquire and localise the surroundings. The simplest way could be to virtually model the environment with a software. For instance Yamamoto [1] allows the modelling of the walls, the doors, the objects, etc, of a building with the real metrics. Their properties like coordinates, orientation and passableness can be defined. This method gives good results but is very long and tedious.

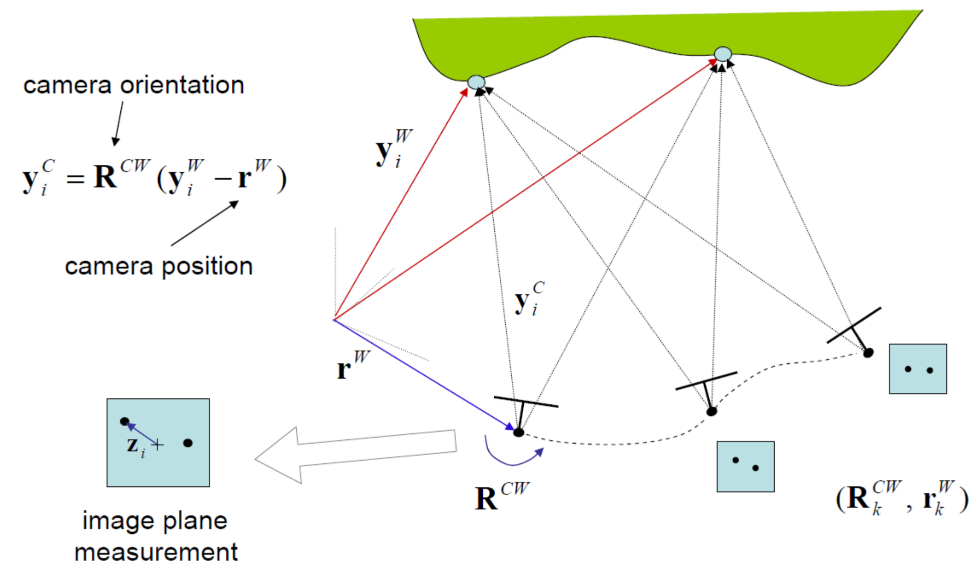


Figure 1: Principle of SLAM (Simultaneous Localisation and Mapping): the position and orientation of the camera change at each acquisition. Some fixed landmarks are considered as references in order to determine the unknown parameters. New landmarks are then mapped. Source: (Davidson et al., 2007)[2].

Another interesting technique is Simultaneous Localisation and Mapping (SLAM), an intelligent on-site learning with one or two video cameras. A possible improvement is to mount the camera on a robot which navigates autonomously in the unknown places of the building. Using the acquired images the system has to localise the camera and at the same time has to map and localise the unknown scanned parts of the environment. The global algorithm is like this:

- Initialization: determination of known reference landmarks.
- Moving of the camera, then detection and matching (usually with a correlation) of the known landmarks in the new image. The measurement operations can be reduced by estimating the movement with a predictor, therefore the search is limited to the uncertainty window.
- Estimation of the actual position and orientation of the camera with the acquired image since the positions of the landmarks are known.
- Detection and localisation of new landmarks using actual and previously estimated positions of the camera. The chosen landmarks are points which are rather easy to determine like edges and corners of shapes. These new points are added to the pool of known landmarks.

The method developed in (Davidson et al., 2007) [2] uses a probabilistic approach to determine the optimal position of the camera at each step. The localisation is represented as a state $\mathbf{x}$ composed of a position vector $\mathbf{r}^W$ and a rotation matrix $\mathbf{R}^{CW}$. All parameters are treated as random variables. C and W are parameters denoting the number of acquisitions and the number of reference landmarks, respectively.

$$\mathbf{q}^{CW} = (q_0^{CW}, q_1^{CW}, q_2^{CW}, q_3^{CW}) \equiv \mathbf{R}^{CW}$$

$$\mathbf{x} = \begin{bmatrix} \mathbf{r}^W \\ \mathbf{q}^{CW} \end{bmatrix}$$

The Fig. 1 represents the parameters in the context. The use of a Kalman filter gives the solution which minimizes the mean square error estimator of the state $\mathbf{x}$ after the acquisitions.

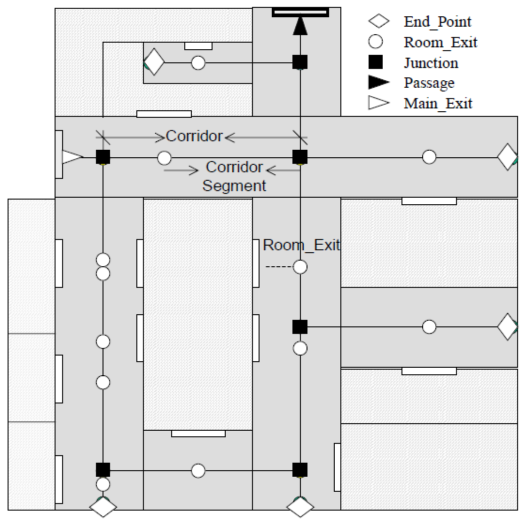


Figure 2: Generation of a graph from a geometrical representation (map). The junctions are the nodes and some other information like room exits are registred on the edges. Source: (Tsetsos et al., 2005) [4].

2.2 Abstract Representation

We now have a *geometrical paradigm* which is very close to the real environment, but it is not usually convenient to use because errors occur during the acquisition, the amount of data is quite large and the exploitation of all this information to find a path is very complex.

To solve the difficulty, a *topological paradigm* is built from the original data. It is an abstract interpretation, usually a graph where a node represents a turn or an intersection (i.e. a decision point) and an edge represents a connection between two points, commonly weighted by the metric distance between these two points [3]. The graph can thus be easily stored and processed by the system during path searching. Geometrical information needed for the guidance has to be stored in the nodes. Moreover, (Tsetsos et al., 2005) [4] advices to integer many landmarks useful to follow the progress on the route like points of interest, exits, obstacles, coffee machines, etc. The Fig. 2 shows how a graph is generated from a geometrical representation.

3 Localisation

One important step in indoor navigation is the acquisition of the position and possibly the orientation of the user. There are lots of different methods for this and in this part only the most significant ones are presented. Orientation may be defined by default as the orientation angle of the line passing through the last two positions.

3.1 Visual Identification

3.1.1 QR-Code

QR-codes (Quick Response-codes) are two-dimensional black-white barcodes which can be placed on strategic places in the building, e.g. on office doors [5]. Each one is associated to specific information (text, hyperlink, etc) and when scanned by a mobile phone it could give the current location of the user to the system. This system may be simple to deploy in some circumstances (e.g. few doors) but is not convenient for proactive guidance.

3.1.2 Visual Analysis

Some technologies allow to visually recognize the user and determine his position in the environment. An interesting actual project is the OpenNI framework based on the Kinect (an Xbox-controller with a video camera and a depth sensor) which analyses the images and provides visual recognition and tracking of many users in real time [6]. An example of the structural recognition of a body from the depth representation is shown in Fig. 3(a).

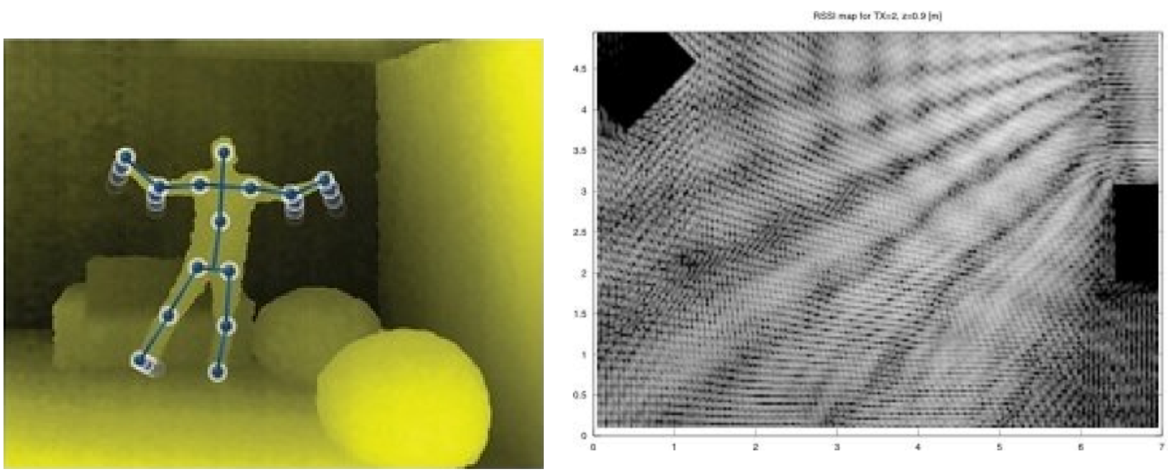
3.2 Proximity Detection

Some devices like active RFID (Radio Frequency IDentification) tags emit a signal all around themselves [1]. If someone passes close to the device, the signal is detected. Since the area of detection is a three-dimensional form close to a sphere around the device, the precision on the localisation may be limited.

3.3 Mapping of Received Signal Strength

This method consists in measuring the received signal strength (RSS) when a signal is emitted. The environment has a specific profile because of the path loss, shadowing by objects and multipath propagation. During the training phase, the system acquires the a priori fingerprint (the map of the power distribution) everywhere in the building (see Fig. 3(b)). Then the system can "recognize" the position of the user from the signal power measured by the receiver. An estimation method like K-Nearest-Neighbour is processed in order to find the best matching. One main limitation of this method is reliability because there are some spatial and temporal instabilities in the modelling, and moreover if the environment changes, for instance people walking or objects moved, the fingerprint alters a lot.

It is usually rather cheap and convenient to use the existing devices in the building, like the Wi-Fi stations [7, 8]. Signals generated outside the building like GSM may also be used; these signals are available almost everywhere, are time-consistent and with the acquisition of many GSM channels the accuracy reaches rather high values [9].



(a) Principle of visual recognition: detection of a human body with the OpenNI framework from a depth image. The typical body structure is recognized. Source: (User guide for OpenNI) [6].

(b) Map of power distribution in a room of 7 by 5 m with the transmitter in the upper right corner. The black shapes are metallic pieces of furniture. The grid resolution measures a few tens of centimeters. Source: (Barsocchi et al., 2010) [8].

Figure 3: Examples of localisation methods.

3.4 Ranges Measurement

This method consists in measuring the time of arrival (TOA) of a signal between an emitter and the user, which leads to know the distance in-between. Since such measurements scale linearly with the propagation distance, a triangulation with at least three different measurements gives the position of the user. The reliability is usually better than with RSS.

A simple method is to emit a specific ultrasound [10] or magnetic signal and to calculate the delay between the emission and the reception. Another solution is to acquire the standard GPS signal. The signal is attenuated by the building but it is still possible to process the signal with a high performing device [11]. Another method consist in generating GNSS-like signals by a WLAN system and use receivers capable of processing GNSS signals (as many smartphones do) [12]. The device will then calculate the position of the user with an accuracy of a few meters.

3.5 Angles Measurement

With this method, angles of incidence of signals are measured. Thus, if for example the user knows the angles of incidence of signals of different emitters, his position can be defined by the point of crossing of the corresponding orientated lines. The user orientation is also quite easy to define. Thanks to signal reflection, standard infrared beacons detect when the user is in front of their own directed beam of infrared light [1]. These sensors are cheap, but the reliability of the measurement is low.

A second method consists in using sensors on one line with a distance d between two consecutive sensors [13]. When an incident wave reaches the line of sensors with an angle θ_l, the difference of path between each consecutive sensor is $d \times sin(\theta_l)$. Thus, the determination of the arrival delay τ_l between two consecutive sensors gives the angle of incidence θ_l.

Another method consists in using passive RFID tags. The RFID reader of the user, when emitting a directional signal, is able to communicate with a RFID tag within an angular range. The orientation of this RFID tag is determined by the angle in the middle of the two communication boundary angles [14].

3.6 Electromechanical Systems

Mechanical methods can also be used, especially in complement to another methods. Inertial sensors implemented in the user terminal, like accelerometers and gyroscopes, allow the integration of the movement [15]. This method is independent from any specific system integrated in the building, but it cannot be used alone because of the increasing imprecision due to the integration. The user orientation can be defined by an integrated magnetic compass.

4 Guidance

4.1 Calculation of the Path

The user has an initial position and wants to reach his final destination. These two points are respectively the *starting point* and the *ending point* of the route to be determined. The possible paths are computed from the graph between the two nodes taken close to these two points and according to the “strength” of each edge. This “strength” depends on the metric distance between the two places represented by the nodes but other parameters may also be taken into account. The document (Lyardet et al., 2008) [5] describes a model for “user centric adaptation” where many parameters are considered:

- Physical capabilities: some places may or may not be accessible depending on the degree of user impairment.
- Location access rights: some places may have restricted access depending on the profile of the user.
- Preferences: attributes of comfort like number of turns, temperature, luminosity, crowdedness.

For the first two points above, some edges in the graph may not be considered in the computation. For the last point, every parameter has to be quantified on a defined scale and influences the “strength” of each edge depending on their importance. A global user-adapted value of preference strength can be computed for a defined path at a specific moment. A method for path finding is to compute all the feasible paths in the graph leading to the destination, then the preference strength of each path is calculated under consideration of importance weights and so the best path is determined. This method is convenient for rather small graphs but for bigger ones a more performing algorithm is needed (e.g. Dijkstra algorithm).

If the user moves away from the route, a new optimal route is computed by considering that the user actual position is the starting point. In some circumstances, even if the user stays on the defined path, it may also be interesting to recompute the optimal path if a parameter changed (closure of a way, changing crowdedness, etc).

4.2 Route Instructions

The abstract path has now to be transformed into comprehensive indications for the user. In other words, semantics has to be generated from information extracted from the graph. It is essentially directional information (e.g. “turn left”) in relationship with landmarks (e.g. “turn left just after the coffee machine”) notified to the user just before a change in direction occurs. These decision points correspond to nodes in the graph. Reported landmarks along the road are also useful because they give a better description of the environment and so help the user in his progress [3].

There are two different representations of the information depending on the point of view on this information:

- *allocentric perspective*: the bird's eye view (also called top-down view) encompasses a large area, e.g. a whole floor of the building, viewed from above without focusing directly on the user. A map which integrates the structures of the environment (generated from the geometrical paradigm for instance), landmarks, the path to the final point with decision points and possibly the position of the user gives a global and contextual knowledge of the whole route. Moreover, this representation is quite useful when there is poor information on the actual position of the user and so it is not possible to focus on his current close environment.
- *egocentric perspective:* this view is user-centred and may be used along with another perspective. For guidance, it can be the succession of instructions as experienced by the user. Each element around the user is localised and represented according to his current position and orientation. The update of instructions is frequently processed and follows the progress of the user on the route. With the user-centred perspective, the user can identify his own perspective with the information he gets: instructions are more adapted, accurate and easy to understand for him.

4.3 Instructions Display

The instructions we now have are abstract data and have to be displayed to the user: they are transformed into multimedia route descriptions based on human senses like the vision, the hearing and even the touch. These instructions have to be compatible with an output device used by the user for the navigation in the building.

The following presents the different ways to display instructions based on human senses:

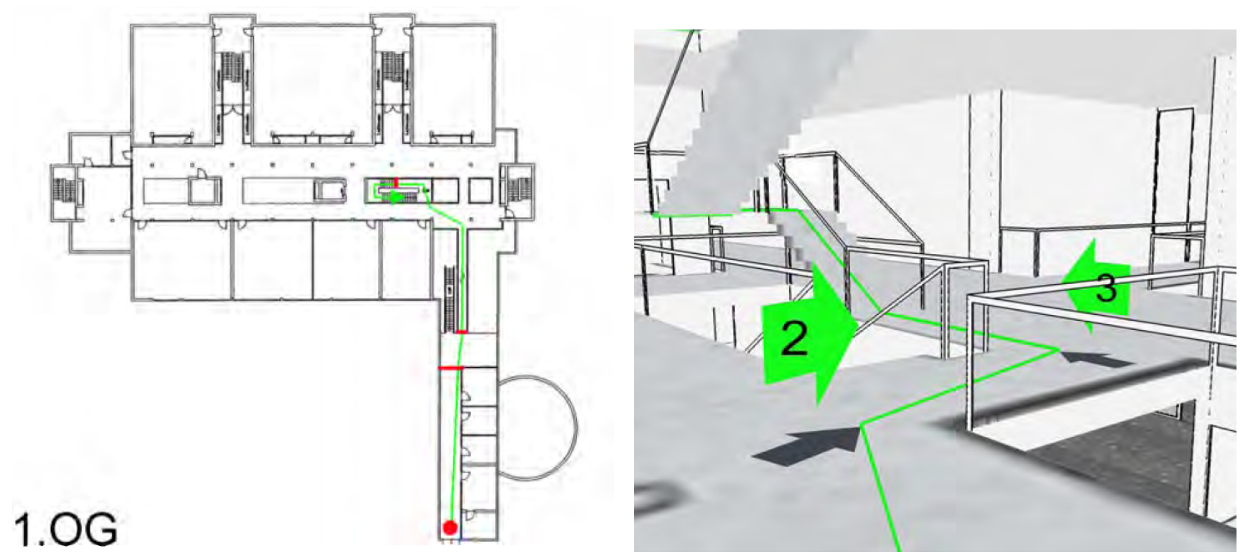
- **Verbal instructions**: a synthetic voice says the instructions like a standard GPS device: "after 10 meters, turn right". One advantage of this method is that the user does not need to maintain his attention on the device.
- **Visual instructions**: multiple views of the scene are possible, depending on several parameters [3]:
  - *distance to the user*: far away to localise globally in the building or close to give nearby context.
  - *inclination*: view from above like a map or with less inclination to give perspective.
  - *orientation*: North-oriented map, user orientation, etc.
  - *dynamics*: static view or view adapted to the user. For the latter case, the view follows the position and the orientation of the user during his movement.
  - *representation of the environment*: pictures of the building, virtual representation, black and white sketch, etc.

The choice of these parameters influences the computational load and the focus attention. Extra representations on the scene has to be added for guidance like arrows showing the direction to follow or texts to emphasize landmarks. Fig. 4(a) represents a sketched top down view with some landmarks (doors and stairs) and the route. Fig. 4(b) represents the equivalent egocentric visual view with the next instructions. Visual instructions can give a lot of complex information but the user has to maintain his visual attention on the screen.

- **Haptic instructions**: it is something less common but it gives good results. The user is guided with tactile objects he has on himself. In the experience of (Bosman et al., 2003) [16], vibration facilities are set on both wrists of experimenters. The information is based in the duration of the vibration: a short vibration on one wrist to give the change in the direction and a long one on both wrists to inform that the direction is wrong. This system is not relevant to give complex information to the user but for basic instructions, it is very intuitive, efficient and rather cheap.

The different ways to give route instructions allow adaptation to user disabilities, for example haptic instructions can be used for blind people (it may also be used to detect nearby obstacles). Moreover information given by several senses may be redundant but generally gives more comfort to the user: a combination of voice and visual indications is more practical because the user needs to watch the visual representations on the screen of his device only when the voice mentions a decision point.

Representation preferences mainly depend on the user. A study reported in (Stahl et al., 2010) [1] shows that the participants of the study preferred visual egocentric route instructions over verbal ones and the preference between an allocentric and egocentric visual view depended on the visual-spatial abilities of the user. Thus it is interesting for the user to have different ways of representation available.



(a) An allocentric view of a route (green line) which diplays the whole floor and some landmarks. The user can anticipate his progress.

(b) An egocentric view of the same route at a decision point. The user has to follow the visual instructions represented by direction arrows.

Figure 4: Examples of visual representations. Source: (Stahl et al., 2010) [1].

5 User Terminal

In the studies encountered, the user has at his disposal a device which gives the navigation information. Many different possibilities exist for this device, mainly depending on the purpose of the navigation. The following is a general discussion about all this.

For a proactive guidance, the user needs a device which is mobile, easy to carry and to use, and which integrates output components like screen and microphone to display the instructions, and possibly a control interface. The personal terminal may also identify the user in order to follow his position and give personalized information, and so allows guiding many users at the same time.

Depending on the purpose and economic considerations (scale of use for example), the guidance in a building can be led by a dedicated device but it may also be interesting to use a connected mobile phone or smartphone. An advantage is that these objects are quite widespread in population. But on the other hand when the terminal is not specific or known, the transmitted information has to be adaptive to make every device compatible.

Under storage space and computational capacity considerations, the database and the processing unit could be installed on the user device (as for standard GPS device) or on one single external server. The latter solution is usually better performing because only the multimedia data have to be transmitted to the user.

A frequent connection has to be established between the system and the device (for the transmission of position or guidance information). A weak connectivity has same effects as weak device computational capacities: limited flow of information. This leads to adapt the display of information to current conditions. For this, the graphical descriptions are structured by a tree in (Butz, 2001) [3]. This allows incremental display of the representations, for example the path is firstly shown, the floor secondly, then more details, etc. This technique adapts the level of the display to the actual resources.

6 Conclusions

In this paper we gave an overview of the main aspects of indoor navigation. This concept can be divided into several main activities which can be treated rather independently: acquisition of the environment, generation of a graph, localisation of user, calculation of path, generation of routing instructions and display of multimedia instructions. The principles and some main techniques were described for all of these activities. Of course other techniques exist but the aim of this work was not to be exhaustive.

A huge amount of research is being done in this area and many new improvements will appear in the future. One can think of a deeper use of smartphones which are now permanently connected, integrate GPS facilities and have higher storage and computational capacities. The actual trends are to better integrate sensorial stimulations for an enhanced immersion, to develop interactions with objects (e.g. augmented reality) and to adapt even more to the user personality in particular with the acquisition of information on the internet.

References

[1] Stahl, C., Münzer, S.: Modeling of multi-level indoor environments: Wayfinding support and empirical studies. In: International Conference on Spatial Cognition. (2010) 2.1, 3.2, 3.5, 4.3, 4

[2] Davison, A., Calway, A., Mayol, W.: Visual SLAM tutorial. Technical report, Imperial College and University of Bristol (2007) 1, 2.1

[3] Butz, A., Baus, J., Krüger, A., Lohse, M.: A hybrid indoor navigation system. IUI'01 (2001) 25–32 2.2, 4.2, 4.3, 5

[4] Tsetsos, V., Anagnostopoulos, C., Kikiras, P., Hasiotis, T., Hadjiefthymiades, S.: A human-centered semantic navigation system for indoor environments. In: International Conference on Pervasive Services, Los Alamitos, CA, USA (2005) 146–155 2, 2.2

[5] Lyardet, F., Szeto, D.W., Aitenbichler, E.: Context-aware indoor navigation. Ambient Intelligence (2008) 290–307 3.1.1, 4.1

[6] The OpenNI Organization: User guide for OpenNI. www.openni.org. 3.1.2, 3(a)

[7] Gallagher, T., Li, B., Dempster, A.G., Rizos, C.: A sector-based campus-wide indoor positioning system. In: International Conference on Indoor Positioning and Indoor Navigation. (2010) 67–68 3.3

[8] Barsocchi, P., Furfari, F., Nepa, P., Potortì, F.: RSSI localization with sensors placed on the user. In: International Conference on Indoor Positioning and Indoor Navigation. (2010) 43–44 3.3, 3(b)

[9] Varshavsky, A., de Lara, E., Hightower, J., LaMarca, Anthony Otsason, V.: GSM indoor localization. Pervasive and Mobile Computing (2007) 698–720 3.3

[10] Filonenko, V., Cullen, C., Carswell, J.D.: Accurate indoor positioning for mobile phones using ultrasound. In: International Conference on Indoor Positioning and Indoor Navigation. (2010) 419–420 3.4

[11] Soloviev, A., Dickman, T.J.: Deeply integrated GPS for indoor navigation. In: International Conference on Indoor Positioning and Indoor Navigation. (2010) 243–244 3.4

[12] Tappero, F., Merminod, B., Ciurana, M.: IEEE 802.11 ranging and multi-lateration for a software-defined positioning receiver. In: International Conference on Indoor Positioning and Indoor Navigation. (2010) 83–84 3.4

[13] Mizutani, K., Ito, T., Sugimoto, M., Hashizume, H.: Fast and accurate ultrasonic 3D localization using the TSat-MUSIC algorithm. In: International Conference on Indoor Positioning and Indoor Navigation. (2010) 423–424 3.5

[14] Uchitomi, N., Inada, A., Fujimoto, M., Wada, T., Mutsuura, K., Okada, H.: Accurate indoor position estimation by the swift-communication range recognition (S-CRR) method in passive RFID systems. In: International Conference on Indoor Positioning and Indoor Navigation. (2010) 313–314 3.5

[15] Edwan, E., Suarez, F., Zhang, J., Loffeld, O.: DCM based attitude estimation using low-cost IMU aided by distributed accelerometers and magnetometers. In: International Conference on Indoor Positioning and Indoor Navigation. (2010) 161–162 3.6

[16] Bosman, S., Groenendaal, B., Findlater, J., Visser, T., de Graaf, M., Markopoulos, P.: Gentleguide: An exploration of haptic output for indoors pedestrian guidance. Lecture Notes in Computer Science (2003) 358–362 4.3

Activity Recognition

Yu Zhang
gu53zot@mytum.de

February 10, 2011

Abstract

Since the 1980s, the activity recognition field has attracted the attention of many researchers. It can strongly provide personalized support for many applications and has connection to many different fields such as medicine, human-computer interaction and sociology. This paper can help people, who have interests in this area but don't quite familiar with it, quickly get the point of the conception of activity recognition, and overview some of the current activity recognition research works. In this article, I will summarize the types of sensors and approaches that be used in activity recognition as an overview. Then I will expound the human motion recognition as a key point, including the general method analysis and an intact example by using motion sensors. At last, there is the summary of this article.

1 Introduction

What is activity recognition? It is an approach that based on a series of observations on the human's actions and the environmental conditions to obtain the recognition of actions and goals. To make this concept easier to understand, a vivid description of Eric Dishman's live can describe it [1]:" Eric Dishman is making a cup of tea and his kitchen knows it. At Intel's Proactive Health Research lab in Hillsboro, OR, tiny sensors monitor the researcher's every move. Radio frequency identification tags and magnetic sensors discreetly affixed to mugs, a tea jar, and a kettle, plus switches that tell when cabinet doors are open or closed, track each tea-making step. A nearby computer makes sense of these signals; if Dishman pauses for too long, video clips on a television prompt him with what to do next."

From this description we can capture some features of activity recognition [2]:" 1) Activity recognition requires sensors, which can generate and receive signals. 2) There should be a software program that can interpret the sensor readings. 3) Activity recognition considers the past. Through learning and inference computer program can predict the future."

With the help of activity recognition, researchers are now able to provide many kinds of support for real-world applications, such as medical assisting technologies, security and environmental monitoring, gaming, sensor-based farming, coal-mine-safety technologies, artificial intelligence research. Several researchers have made effort on this. For example, by using automatically monitoring human activities home-based rehabilitation can be provided for people suffering from traumatic brain injuries [3]; Activity recognition can also predict transportation modes [4]; Abnormal human activity can be detected by activity recognition for security monitoring [5].

2 Overview

The area of activity recognition covers various aspects, including hardware related and software related aspects. To get the point as soon as possible, I will respectively expound her in two parts: types of sensors used in activity recognition and approaches of activity recognition.

2.1 Types of Sensors and Sensory Devices

Sensors open a window to the outside world for computer systems. Building an activity recognition system requires a mount of original data from sensors and sensory devices. Different sensors are suitable for different kinds of activity recognition missions. RFID, GPS, WLAN, mobile phones, cameras and motion sensors will be introduced below.

Radio-frequency identification (RFID) uses an integrated circuit for storing and sending radio-frequency (RF) signal. The RFID tags cover several meters. It has been widely used in product tracking and identification, especially in logistics operations.

The Global Positioning System (GPS) [6] is a global system based on satellites that send RF signals. Receivers can determine their current locations in an outdoor environment. Based on the location sequences high-level inference can be done to determine transportation modes, goals and intentions [7].

The Wireless Local Area Network (WLAN) is an IEEE 802.11a/b/g/u/f wireless network in the 2.4GHz (802.11b/g/u) or 5.9GHz (802.11a/f) frequency bandwidth used in an indoor or even outdoor area. WLAN devices are especially useful for locating a user and tracking human's movement in indoor and outdoor environment where GPS is not suitable.

Mobile phones nowadays have advanced sensors, such as gyroscopes, accelerometers, and compasses, for measuring some predefined activity of mobile users. A common feature of these sensors is that they are cheap and very widely available in our everyday lives.

Cameras are also selected by many researchers, who have interests in field of activity recognition such as objects recognition, human motion recognition and human tracking problems. Kinect is a good example. Kinect for Xbox 360 or simply Kinect is a new video game system that Mirosoft recently developed. The technology of Kinect is mainly based on vision-based activity recognition (human motion recognition). It brings a new revolution to the concept of video game, because it lets users control games without a handhold game controller. The main technology is vision based activity recognition. Using "RGB camera, depth sensor and multi-array microphone running proprietary software" [8] provides Kinect full-body 3D motion capture, facial recognition, and voice recognition capabilities. It guides users to a Xbox live, that living room becomes smarter [9]. It responses to your gestures and listens your command. You can watch a movie without a remote, play a game without controller. It helps you to connect with friends and entertainment that you care about. Find counting on your TV is so simple, that everyone can use it.

The main principle of motion sensors are based on their physical features. They can sense some specified characteristics from the environment, i.e. pressure, temperature, tilt. Because of many advantages such as the low price and portability, motion sensors are widely used in activity recognition field to predict the locations, trajectories, actions etc. Following three kinds of popular motion sensors are introduced: Accelerometers, Ball switches and capacitive sensor.

An accelerometer measures proper acceleration relative to free fall. There are single- and multi-axis models. It is possible to detect magnitude and direction of the acceleration as a vector quantity. It can be used to sense orientation, acceleration, vibration shock, and falling. Accelerometers recently play an important role in personal digital products, i.e. smart phones, digital cameras and laptops. For example, an LIS302DL accelerometer [10] is used in iPhone and iPod touch to detect tilting situation of the device. Nowadays many digital cameras contain accelerometers. It is possible to know how to orientate photos and rotate the current photo when viewing. For device integrity, accelerometers are also used in laptops to detect dropping for hard disk protecting. Lenovo's 'Active Protection System', Apple's 'Sudden Motion Sensor' and HP's '3D Drive Guard' [10] all use this technology.

Ball switches can detect the tilt, upside down, rotation, or slight vibration of the objects. It's simple ON and OFF signals make roll ball switches easy for design. Because it is simple, reliable and inexpensive, it has a wide range of applications, for example, security industry, wireless industry, anti-theft industry and entertainment industry (i.e. Fit in Toy Designs) and Energy Saving.

Capacitive sensing is recently announced method for human-computer interaction. Because of capacitive sensor's unique capacitive features and limitations make it as a great option to optical and ultrasonic object tracking. It can detect and track conductive as well as non-conductive objects. Wimmer. et *al*. [11] have done much works on this area. They presented a toolkit for realizing capacitive sensing applications for human-computer interaction in pervasive computing systems by using a self developed CapToolKit. This toolkit can rapidly realize prototypes and systems that are able to detect the presence of humans and objects.

As above I have introduced the basic hardware composition (sensors) for activity recognition. In the following part I will introduce the realization approaches used in activity recognition in detail, such as Machine learning approach and data mining based approach.

2.2 Approaches of Activity Recognition

2.2.1 Machine Learning Approaches

Machine learning is an approach that uses empirical data (from sensor data or databases) to reckon behaviors. It cares about the design and development of algorithms. It has two major tasks: learning to recognize complex patterns and making intelligent decisions based on data. Following logic reasoning and probabilistic reasoning are introduced as two machine learning approaches that are used often by researchers to solve activity recognition problems.

Logic reasoning records all logically consistent explanations of the observed actions. As a result, all possible and consistent plans or goals must be considered. A serious problem of logic reasoning is the inability or inherent infeasibility to represent uncertainty. Probabilistic reasoning becomes the dominate approach for activity recognition. Probability theory and statistical learning models are well suited to reason about actions, plans and goals. For example, Intel Research (Seattle) Lab and University of Washington have done some important works on using sensors to detect human plans [12]. In general, there are two main types of models: discriminative models and generative models.

Discriminative models directly represent the conditional distributions of the hidden labels, which is given all the observations [13]. Logistic regression and conditional random fields are both typical training discriminative models. During the parameters the system is able to adapt to maximize the conditional likelihood of the data. Discriminative models are well suited for tasks that include complex and overlapped evidences. In contrast, generative models represent the joint distributions and use the Bayes rule, such as naive Bayesian models, hidden Markov models, and dynamic Bayesian networks, to obtain the conditional distribution [13].

2.2.2 Data Mining Based Approaches

Different from traditional machine learning approaches, Dr. Gu et *al*. [14] proposed a new approach based on data mining. With this method, the problem of activity recognition is considered as a pattern-based classification problem. This approach is based on discriminative patterns. And it describes obvious changes between any two activity classes of data. It can recognize sequential, interleaved and concurrent activities in a unified solution. There are many advantages of data mining approach, for example, training is not needed for interleaved and concurrent activities. In addition, it will not include noise patterns. So the result can tolerant more noise and the noise is random distributed among classes.

3 Human Motion Recognition

3.1 Human Motion Analysis

The human motion analysis can be separated into three parts [15] (see Fig. 1): Body Structure Analysis, Tracking and Recognition. A general method of human motion is shown as the picture. When analyzing a human motion, the first thing should do is to analyze the body structure of a person. We can establish a model to describe human's structure, or never use a model to present this. Then the second step is tracking. We can use a single camera or multiple cameras to take images and extract the features of image sequences. Based on successful tracking, the correct rate of human motion recognition will be increased. State-space and template matching are two kinds of method usually used for recognition.

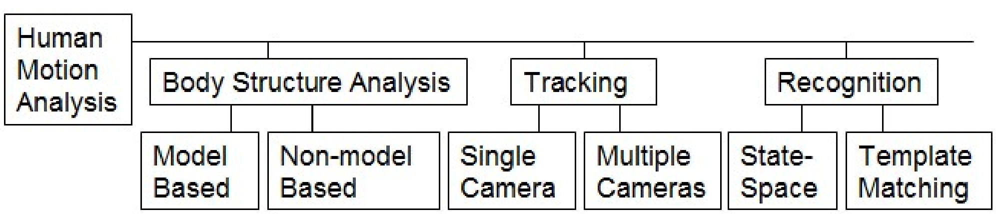


Figure 1: A general method for human motion analysis consists of three steps: Body Structure Analysis, Tracking and Recognition. [15]

3.2 Approaches of Human Body Parts Analysis

The approaches of human body parts analysis can be separated into non-model based and model based approaches. The common features of both methodologies are that they follow the general structure: feature extraction, feature correspondence and high-level processing. In a rut, human bodies are represented as stick figures, 2D contours, or volumetric models. As a result, body segments can be represented as lines, 2D ribbons, and elliptical cylinders, respectively [15].

3.2.1 Non-model Based Approaches

The non-model based approach doesn't consider the shape of object. If we don't consider a priori shape models, heuristic assumptions are usually used to set up the correspondence of joints in an image sequence. The simplest method to represent a human body is the stick figure. The motion of joints is the key to estimate motion and recognize the whole figure. Another method to describe the human body is 2D contours. Under this assumption, the human body segments are analogous to 2D ribbons or blobs.

3.2.2 Model Based Approaches

The model-based approach uses the information about the object shape. Because human eye usually uses a priori shape models to explain moving figures, most methods for motion analysis use models to learn from previous images. It is usually difficult to establish feature correspondence between consecutive frames compared with the non-model based approach. However, it provides the possibility to achieve high-level tasks at the final stage.

3.3 Approaches of Tracking Human Motion

The aim of tracking is to set up correspondence between consecutive frames. They are based on features such as position, velocity, shape, and color. Typically, the tracking process is about matching between images by using pixels, points, lines and blobs. They are based on the motion, shape and other visual information. There are two general correspondence models: iconic models and structural models. The first model uses correlation templates and the second model uses image features. Following we discuss the problem of tracking human motion in two scenarios: with a single camera and with multiple cameras.

3.3.1 Single Camera Tracking

Under this scenario, images are taken from one single camera. And most methods for tracking moving humans use this scenario. Usually points and motion blobs are the features used for tracking. Low-level feature such as points are easier to extract but relatively more difficult to track than higher-level features such as lines, blobs and polygons. So there is always a tradeoff between tracking efficiency and features complexity. Another commonly used feature for tracking is 2D blobs or meshes. The disadvantage of using only one camera is that the view captured by the camera is relatively narrow.

3.3.2 Multiple Camera Tracking

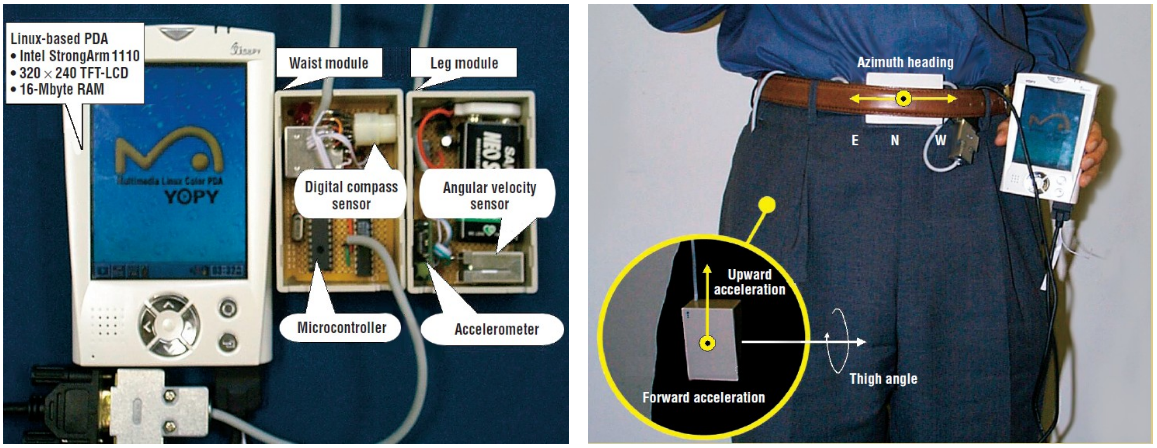
A multiple camera scenario helps to solve the problems that a single camera has. If the subject disappears from the view of one camera, it will appear in the view of another camera in the system. For example, Shiloh L. Dockstader and A. Murat Tekalp [16] set up a distributed multiple cameras, real-time computing platform for tracking multiple interacting persons in motion. The disadvantage of this strategy is that images taken from multiple perspectives must correspond to the same spatial coordinate. Meanwhile camera placement is also very important. Robert Bodor [17] focuses much on this problem, trying to determine how to place cameras to provide image input to optimize the system's ability.

3.4 Approaches of Human Activity Recognition

For human activity or behavior recognition most works have been done based on using template matching approaches and state-space approaches.

3.4.1 Template Matching Approaches

Template matching approaches compare the feature extracted from the given image sequence to the patterns stored before during the recognition process [15]. The advantage of using template matching approach is its inexpensive computational cost. However, it is sensitive to the duration of motion.



(a) The proposed PDA and two sensing modules: leg module and waist module

(b) The use of sensing modules and the direction of each measurement

Figure 2: The hardware components and the use of the wearable system. [18]

3.4.2 State-Space Approaches

State-space approaches are suitable to understand the human motion sequence. It is widely used to predict, estimate, and detect signals over various applications. The features used in this technique are points, lines and 2D blobs. And one typical model used in this approach is the Hidden Markov Model (HMM). The advantage of this approach is that the duration of motion is no longer a problem because each state can repeatedly visit itself. However, because this method usually needs intrinsic nonlinear models, it requires complex computing interactions to seek a global optimum in the training process.

3.5 Motion Sensor-based Human Motion Recognition

Besides using vision-based sensors to recognize the human motion, motion sensors (i.e. acceleration sensors) are also widely used to extract original data for human motion analysis. Majority of motion sensors are inexpensive and with small volume, so that they can be easily wear by people. Here I will expound a paper in detail, that Lee et *al*. [18] showed how to recognize human activity and location by using wearable sensors. The aim of this paper cares two tasks: The first task is to recognize a person's behavior for example sitting, standing and three types of walking behaviors - walking on level ground, going up a stairway and going down a stairway. The second task is to determine a person's current location by detecting transitions between preselected locations. My focus is on the first task.

As following I will introduce Lee's et *al*. [18] hardware system for motion recognition. The system consists of two parts: a Linux-based PDA and a sensing module. The body-worn sensing module is carried out in two separate 50*35*75mm boxes: leg module and waist model (see Fig.2 (a)). The use of these two modules are shown in Fig.2 (b). One box the leg module contains the biaxial accelerometer and the gyroscope. It is put in the user's left or right trouser pocket in order to measure the acceleration and angle of the user's thigh. The other box the waist module is located at the middle of the user's waist to detect direction when the person moves. In general, The sensing block reads the data from the sensors, then sends the data to PDA via its serial port. The PDA implements a set of preprocessing tasks such as filtering and computing statistical properties. Finally, using this features the unit motion recognizer identifies the type of a unit motion, which is predefined from one of the five unit motions.

As below shown, there is the presentation of the steps followed by Lee et *al*. [18]. At first, the system requires a set of data for training to determine the parameters of the unit motion recognizer. The parameters collected from sensors and some new defined basic feature values are showing as below.

1. Data collection. We use 3D position vector as a descriptor. The data collected by the sensors are:

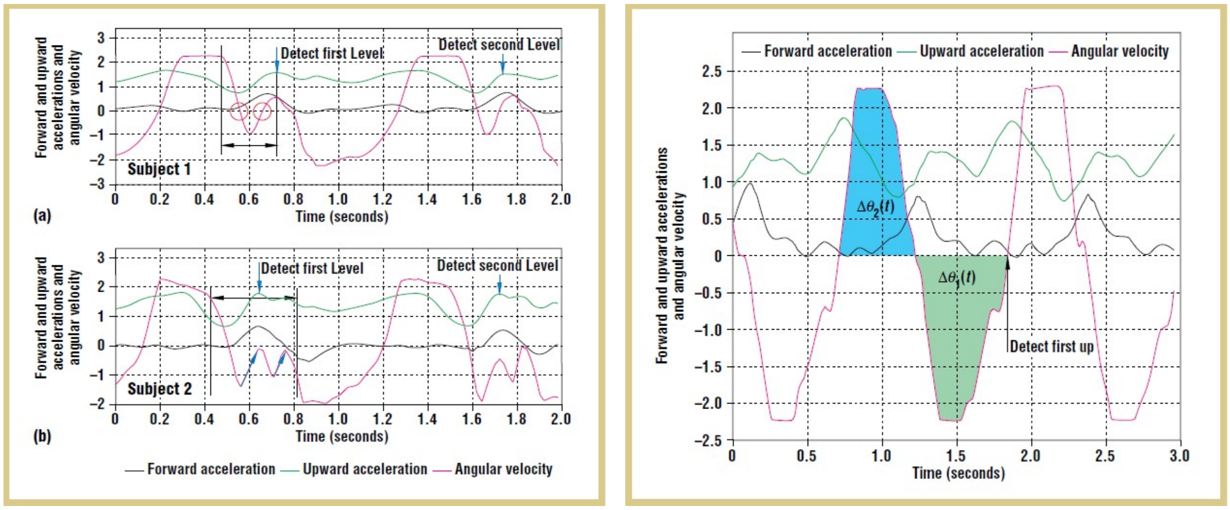
   - Accelerometer in the leg module measures forward and upward accelerations of user's thigh, which is donated $a_x(t)$ and $a_z(t)$.

- A digital gyroscope obtains angular velocity $\dot{\theta}(t)$. Using a digital integrator the system measures the angle $\theta(t)$ of the user's thigh movement.
- The digital compass sensor gives one of azimuth headings (N,E,S,W) as logical signal every 20 milliseconds, and then it sends the data to the PDA through a serial communication channel.

2. Basic features definition. These values are selected to predict the unit motions, which are important for a robust and reliable recognition. They are as following:
   - A standard deviation (over 50 samples) of forward acceleration $\sigma_x(t)$, upward acceleration $\sigma_z(t)$, and thigh angle $\sigma_\theta(t)$.
   - Three angle differences $\varDelta\theta_1(t)$,$\varDelta\theta_2(t)$,$\varDelta\theta_3(t)$. Each of the value can be obtained through integration of angular velocity in a time interval between zero crossing of $\theta(t)$.
3. Motion recognition. The unit motion recognition process is started using feature values above. In general, it can be separated into two problems, non-walking walking behavior recognition and walking behavior recognition.
   (a) Non-walking behaviors recognition. By using the accelerometer to detect an absolute gravitational acceleration we can easily recognize sitting and standing behaviors.
      - Sitting condition: $\sigma_\theta(t) > 16$, $\varDelta\theta_1(t) > 70°$, $a_x(t) > 0.7g$;
      - Standing condition: $\sigma_\theta(t) > 16$, $\varDelta\theta_1(t) < -70°$, $a_x(t) < 0.3g$.
        Here, g presents one gravitational acceleration.

   (b) Walking behaviors recognition. There are three types of walking: a level walking, an upward walking and a download walking. The system can not only recognize the user's activities but also able to count the number of steps. So the system must know how to distinguish "one cycle of human walking" (called the "gait cycle") [18]. It is defined in Lee's paper as: "in terms of a time interval during which one sequence of a regularly recurring succession of events is completed." [18] Determining a walking behavior must in terms of truth values of each proposition $\omega^i, i = S, N, F, U, D$, which are resulted by calculating the input vectors through a simple fuzzy-logic reasoning method. This mathematical methodology is a human based learning processes. Compared with machine learning algorithms it is more suitable for representing what the human has learned. For walking behavior recognition, there are 4 steps as shown below.
      - Test level walking conditions & calculate the truth values of three subcategories: slow, normal or fast walking behavior. With classification, it surly helps to improve the performance of the proposed location recognition method.
        At first the system tests following conditions to make sure that the current behavior is a level walking behavior. It finds the positive peak value of upward acceleration $a_z(t)$ (the blue down arrows in Fig.3 (a)-(a)). Then, it tests the following conditions:
        – $\sigma_x(t) > Th_{\sigma_x}$ AND $\sigma_z(t) > Th_{\sigma_z}$ AND $\sigma_\theta(t) > Th_{\sigma_\theta}$, where $Th_{\sigma_{x,z,\theta}}$ are threshold values for three feature values.
        – Wether the number at the zero crossing of $\dot{\theta}(t)$ in interval is ≥ 2 (the red circles in Fig.3 (a)-(a)). If it is smaller than two, then test wether the number of angle changes is ≥ 2 (the blue upward arrows in Fig.3 (a)-(b)).

        If both conditions are satisfied, then the current behavior can be recognized as a level walking, and the system executes the truth value of three subcategories: slow, normal and fast walking. The input vector set is: $\vec{u}_L(t) = \{u_1, u_2, u_3\} \equiv \{\sigma_x(t), \sigma_z(t), \sigma_\theta(t)\}$. Via a simple fuzzy-logic reasoning method, as a result, the truth values of each proposition are as:
        $\omega_l^S = min(\mu_1^S(u_1), \mu_2^S(u_2), \mu_3^S(u_3))$,
        $\omega_l^N = min(\mu_1^N(u_1), \mu_2^N(u_2), \mu_3^N(u_3))$,
        $\omega_l^F = min(\mu_1^F(u_1), \mu_2^F(u_2), \mu_3^F(u_3))$.
        Here min operation is used as the AND operation in the fuzzy rules.
      - Calculate the truth value of upward proposition.
        The upward walking behavior recognition is also based on the fuzzy logic reasoning method, and separated into two steps.
        – The system try to find the end of a cycle of up behavior. It can be found when the angular velocity goes to positive near the moment of positive peak $a_x(t)$ (Fig.3 (b)).



(a) Sensor signals for level walking behavior for two subjects (b) Sensor signals for upward walking behavior

Figure 3: Typical trajectories of sensor signals for level and up walking behavior [18]

- The system performs the same fuzzy-logic reasoning process. The input vector are defined as $\vec{u}_U(t) = \{u_1, ..., u_5\} \equiv \{\sigma_x(t), \sigma_z(t), \sigma_\theta(t), \Delta\theta_1(t), \Delta\theta_2(t)\}$. As a result, the truth values of each proposition are as $\omega^U = min(\mu_1^U(u_1), ..., \mu_5^U(u_5))$.

- Calculate the truth value of downward proposition.
The downward walking recognition uses the same method as upward behavior recognition but with a different set of input values: $\vec{u}_D(t) = \{u_1, u_2, u_3\} \equiv \{\Delta\theta_1(t), \Delta\theta_2(t), \Delta\theta_3(t)\}$. This reasoning is performed whenever zero crossing of the angular velocity occurs. As a result, we can get the truth value as: $\omega^D = min(\mu_1^D(u_1), \mu_2^D(u_2), \mu_3^D(u_3))$.
- Definite the walking behavior.
Finally, the unit motion recognizer finds a maxim value from the obtained truth values $\omega^i, i = S, N, F, U, D$. S,N,F are all recognized as a level walking; U is recognized current motion as a upward walking and D is recognized as a downward walking. If $Max(\omega^i) > Th_f$, the recognizer eventually determines the current step as one of the walking behaviors. Otherwise, it is considered as a missing step.

As the experiment result showed, the level walking behavior recognition obtains a better result than upward and downward walking behavior recognition. In total 978 level steps just 2.95% of steps can not be recognized and there are very small possibility (more or less than 0.6%) to recognize a level walking as a upward or downward walking behavior. In contrast, upward has 5.65% ratlos of unit motion recognizer in total 195 steps and downward has 6.63% of recognition ratlos in total 199 steps.

4 Summary

I give an overview about past developments of activity recognition including two parts: sensor and sensory devices (hardware related) and approaches such as machine reasoning and data mining based reasoning (software related). Different types of current widely used sensors are explained in details including their properties, suitable areas, applications, advantages and disadvantages. Popular approaches used in activity recognition field are also introduced including their features, classifications, algorithms and suitable fields. The human motion recognition is as a key point to illustrate. At first a general method for human motion analysis is introduced, which is inclusive three parts, motion analysis of the human parts, high-level tracking and recognition of human movements. After that an example of using accelerometer to recognize human unit motion, such as sitting, standing and walking, is expounded in details. The experiment shows good result by using this method.

References

[1] Huang, G.: Monitoring Mom: As population matures, so do assisted-living technologies. Technical Review **20** (2003) 1

[2] Yang, Q.: Activity recognition: linking low-level sensors to high-level intelligence. Proc. IJCAI-09 (2009) 20–26 1

[3] Pollack, M., Brown, L., Colbry, D., McCarthy, C., Orosz, C., Peintner, B., Ramakrishnan, S., Tsamardinos, I.: Autominder: An intelligent cognitive orthotic system for people with memory impairment. Robotics and Autonomous Systems **44**(3-4) (2003) 273–282 1

[4] Liao, L., Patterson, D., Fox, D., Kautz, H.: Learning and inferring transportation routines. Artificial Intelligence **171**(5-6) (2007) 311–331 1

[5] Yin, J., Chai, X., Yang, Q.: High-level goal recognition in a wireless LAN. In: PROCEEDINGS OF THE NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE, Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999 (2004) 578–584 1

[6] Pan, S., Kwok, J., Yang, Q., Pan, J.: Adaptive localization in a dynamic wifi environment through multi-view learning. In: PROCEEDINGS OF THE NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE. Volume 22., Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999 (2007) 1108 2.1

[7] Zheng, Y., Liu, L., Wang, L., Xie, X.: Learning transportation mode from raw gps data for geographic applications on the web. In: Proceeding of the 17th international conference on World Wide Web, ACM (2008) 247–256 2.1

[8] Wikipedia: Kinect. `http://en.wikipedia.org/wiki/Kinect` (Dec 2010) 2.1

[9] YouTube: Microsoft e3 kinect video kinect demo. `http://www.youtube.com/watch?v=LUsiFjeRWjw` (Dec 2010) 2.1

[10] Wikipedia: Accelerometer. `http://en.wikipedia.org/wiki/Accelerometer` (Dec 2010) 2.1

[11] Wimmer, R., Kranz, M., Boring, S., Schmidt, A.: A capacitive sensing toolkit for pervasive activity detection and recognition. In: Pervasive Computing and Communications, 2007. PerCom'07. Fifth Annual IEEE International Conference on, IEEE (2007) 171–180 2.1

[12] Perkowitz, M., Philipose, M., Fishkin, K., Patterson, D.: Mining models of human activities from the web. In: Proceedings of the 13th international conference on World Wide Web, ACM (2004) 573–582 2.2.1

[13] Liao, L.: Location-based activity recognition. PhD thesis, Citeseer (2006) 2.2.1

[14] Gu, T., Wu, Z., Tao, X., Pung, H., Lu, J.: epsicar: An emerging patterns based approach to sequential, interleaved and concurrent activity recognition. In: Pervasive Computing and Communications, 2009. PerCom 2009. IEEE International Conference on, IEEE (2009) 1–9 2.2.2

[15] Aggarwal, J., Cai, Q.: Human motion analysis: A review. In: Nonrigid and Articulated Motion Workshop, 1997. Proceedings., IEEE, IEEE (2002) 90–102 3.1, 1, 3.2, 3.4.1

[16] Dockstader, S., Tekalp, A.: Multiple camera tracking of interacting and occluded human motion. Proceedings of the IEEE **89**(10) (2002) 1441–1455 3.3.2

[17] Bodor, R.: Multi-camera human activity recognition in unconstrained indoor and outdoor environments. PhD thesis, University of Minnesota (2005) 3.3.2

[18] Lee, S., Mase, K.: Activity and location recognition using wearable sensors. Pervasive Computing, IEEE **1**(3) (2002) 24–32 2, 3.5, 3.5, 3b, 3

Ubiquitous Campus

Wei Chen
chenwei824@mytum.de

February 10, 2011

Abstract

We are living through an era full of technological innovations, that have changed and are changing the campus sustainably during the last decade. This paper highlights how ubiquitous campuses are coming into existence all over the world.

1 Introduction

Since computer and internet have won every man over, it has even become impossible for us to imagine about how life would be, if we had to do without them. Ten years ago Bob, who had not finished his studies of informatics yet, possessed a PC with an CRT Display, that was like a lead weight and he used his 56k analogue modem in order to surf the web. He beamed with pride and joy whenever he swaggered about his IBM 586. Today Leon, who has just begun his studies at TU Munich, would just smile benignly at Bob. Nowadays his mobile device, which is equipped with a three-inch touch screen, is online day and night. Terms like ISDN, cables and unattainability are buried in oblivion. Today Leon speaks about online services, web portals, mobile solutions, UMTS, location based services (abbreviated as LBS) and so forth.

Hardly a month passes without some breathtaking development in science and technology and life on campus has become so different. Twenty years ago while Bob was sitting in the lecture hall and writing off from the black board continuously, the professor tried to utilize every blank area of the blackboard in hope that he does not need to wipe the it again. How time flies! Today Leon can settle back after he has entered the lecture room. He does not need to write off anything any more. What is writing off still good for when the professor's notes will be provided as PDFs in online portals few hours later? Bob used to searched through many index cards in order to find a book in the university's library, he payed in cash for the lunch in the canteen and he had one key of each of the laboratories where experiments were carried out. Today electronic online databases have replaced index cards, student card has replaced cash and keys. Maybe we are going to have an all-round solution tomorrow, that is called smartphone, a kind of mobile device.

Ubiquitous computing have gained in importance. This paper presents why and how it has become so important on campus and what kind of projects in respect of ubiquitous campus have already been launched.

2 Requirements and Hitherto Pilot Experiments

In the 21st century the technology revolution will move into the everyday, the small and the invisible.[1]

Mark Weiser

2.1 Requirements

The requirements of a ubiquitous system have been classified in three categories: system, software, business and organization. System refers to the platform i.e. the hardware of the system or the stem running the subject software on top of it. The term software represents the services and the components of system that have been derived from the ubiquitous system. Business refers to the actors who are providing the components and services for the development of the system. Organization includes development methods and the processes needed for the development and integration of the system services. [3] This paragraph tries to answer this question by specifying the system requirements of ubiquitous computing on campus and pointing up its hitherto existing stages.

2.1.1 System Requirements

1. The **Digital revolution**, that consists of three eras, made ubiquitous computing possible. While the first era concern mainly big enterprises, the second era layed the foundation stone of ubiquitous computing, for integrated chips became smaller and cheaper. We now live in the third era of the digital revolution, that is also called the era of ubiquitous computing and is branded by smart and small computers that are pervading every area of our daily life.

2. **Affordable mobile devices** are also an essential requirement for an ubiquitous campus. It can be a notebook or even better for practical reasons a smartphone that is equipped with reasonably good computing power. For educational using the purpose of mobile devices are to [5]

   - promote "anywhere, anytime" communication
   - organize teaching and learning in and out of the classroom
   - promote collaborative learning among students
   - provide students with an optimal learning experience and technological skill

   In the present time, mobile devices are becoming an important tool with everyday utility, so much so that for some students it is the only mode of communication with other people.

3. **A campus wide wireless local area computer network** plays a key role to ubiquitous campus. This kind of infrastructure enables the possibility to provide not only the access to the world wide web but also location based services. In order to use their smartphones in its entirety, students do not need to use the expensive UMTS connection anymore. Mark Weiser, the father of ubiquitous computing said once: Wireless Communication will be key to tying together the diverse functions of the many computers around us, a compact campus environment makes that easier. [2]

2.1.2 Software Requirements

1. More and more **Location Based Services** are provided. Since smartphones have becoming widely used, various sensors integrated on a mobile phone can be used to get location information. Knowing student's location helps the university to offer more relevant services based on surrounding environment. For LBS is relatively new and few universities have introduced this kind of service till now, I would like to discuss about it in a separate subchapter later.

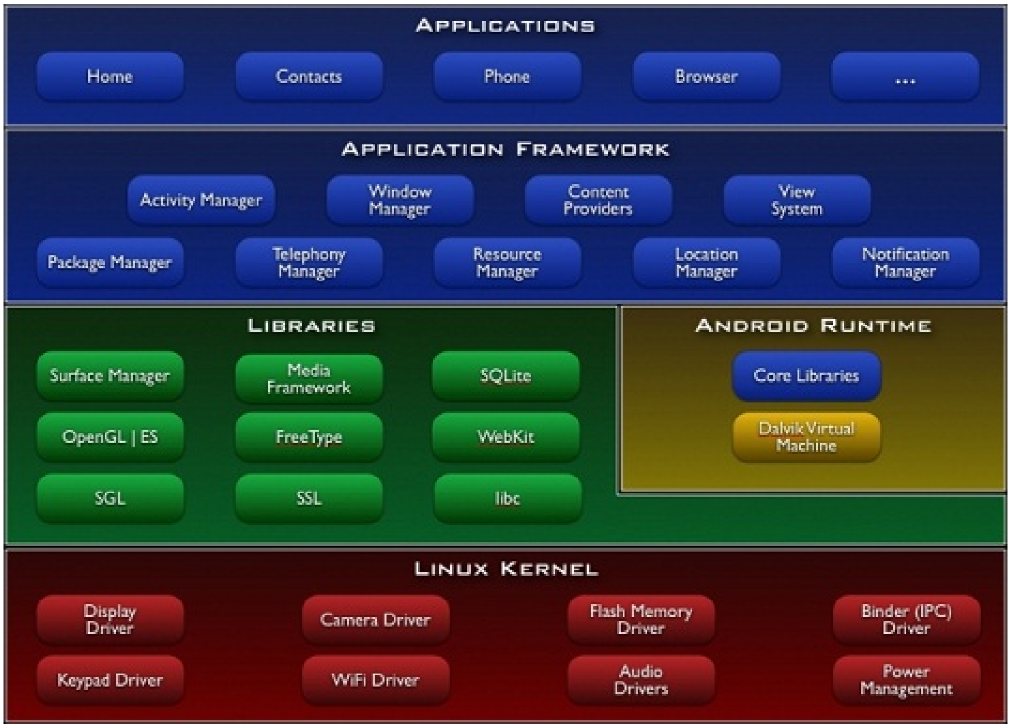


Figure 1: Android Operating System Architecture Source: http://www.anddev.org/images/android/system_architecture.png

2. An software operation system has to be characterized by it's **interoperability, heterogeneity, mobility, survivability and security**. Google Android is one of the efforts currently being carried out by Google in order to deliver a ubiquitous system on mobile phones. [4] Google Android is interoperable thanks to its content provider that is responsible of handling the published data. Both contacts data and course time schedule can be shared easily. It is also heterogeneous and mobile by making use of libraries like OpenGL and SQLite. The most important component Android runtime, that is an optimized version of java runtime for mobile phones, make porting applications to other machines and devices child's play because Android applications on upper layer won't be using the direct interfaces below. Since its architecture is based on LINUX, it adopts and benefits from all the features a linux kernel provides. SSL is also available on Android so that data can be interchanged encrypted. (see Fig. 1) Another very important aspect is that google Android is free and open source, and this has a positive affect on the the price of a mobile device with Android as OS.

2.1.3 Business And Organization Requirements

In this paper, we only regard universities and colleges as actors and the providers for the components and services of ubiquitous computing. For universities and colleges often are willing to explore opportunities of wireless location-aware computing in the university setting, they often have installed the essential hardware requirements as pioneers. Thus university campuses have become favourable area to start deploying and test the newest ubiquitous computing technologies.

2.2 Pilot Experiments At Very Beginning

This and the next subchapters give an overview of universities worldwide, where pilot experiments to the ubiquitous computing on campus were run or are being launched. At the beginning of the 21th century projects were launched in order promote and fulfill the requirements of ubiquitous computing, but this subchapter does not deal with these project, but it deals with projects that were launched once the hardware system environment was available and fully functional.

2.2.1 University-wide course management tool and other online portals

Course management tools can facilitate a student's and a teacher's life immense. And since ubiquitous computing on campus has become a commonly used term, many university-wide course management tool projects have already been introduced successfully. One of the biggest is **K-tai campus**[7] , that is developed by the National Institute of Multimedia Education in Japan for various universities and colleges in Japan. In Comparison with many European and American universities who prefer to have their own individualized course management system, for example: TUMOnline, K-Tai was reduced to a common denominator, so that it can be used by every college and university in Japan. Its goal was to create a common online course management tool. K-tai was developed to support both the mobile and the desktop platform, although it is designed from the view that teachers mainly use PCs and students more often use mobile phones. Users were classify into five user access categories: administrator, teacher, staff, student and visitor. Every user had their own user access depend on their status.

K-tai uses a very simple server-client architecture. The Server consists of a Redhat Linux Server with PostgreSQL as a database and an Apache Client, which also supports PHP. The Client can be any internet Browser.

2.2.2 Future libraries

Bit by bit printed books will disappear since they can be replaced by ebooks. No more buildings will be needed to store books, and no more employees will be needed to clean out the libraries. The only equipments we need are a computer to open the ebooks and a database with a big storage where ebooks are stored centrally. The virtual library for the future consists of ebooks that can be found very easily online and downloaded by every student.

The free zone, which has arisen from book-racks that have been cleared away, can be turned into learning rooms, where students can spend the time together.[5]

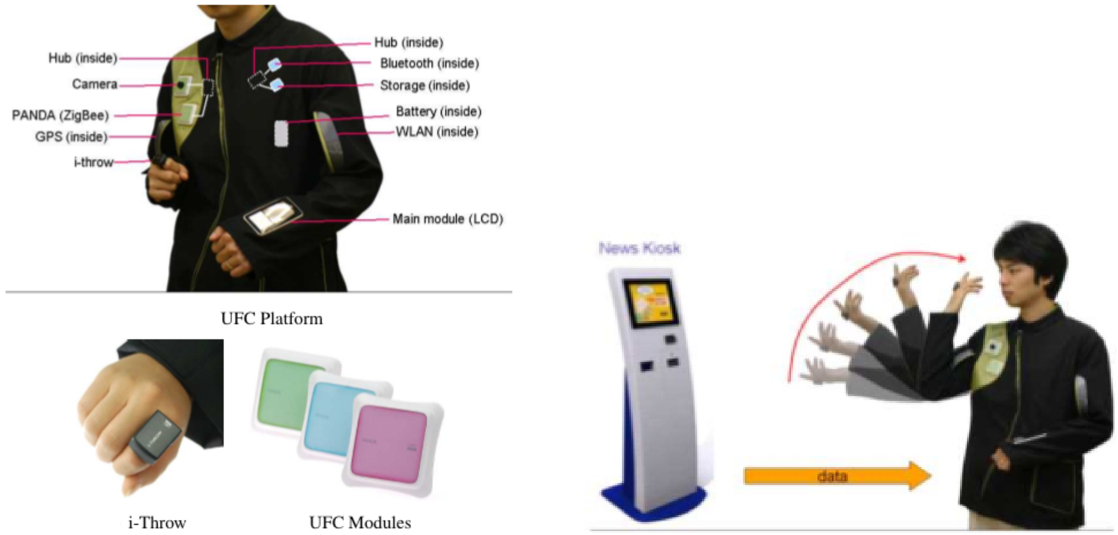
2.3 Development Trend - Location Based Services

Ubiquitous computing names the third wave in computing, just now beginning. First were mainframes, each shared by lots of people. Now we are in the personal computing era, person and machine staring uneasily at each other across the desktop. Mark Weiser said, "Next comes ubiquitous computing, or the age of calm technology, when technology recedes into the background of our lives."

As already mentioned in the previous chapter, location based services have a big potential on campus the next years. Especially by means of new technologies like HTML5 and 3rd party tools like Google Maps new opportunities to the ubiquitous campus are given. Staff work time logging can be handled more precisely and easily, and points of interest and also advertisements can be offered more selectively. Location based campus social networks are getting formed and people can be connected to each other on a new way. This subchapter specifies some opportunities of location based services on campus.

2.3.1 U-TOPIA Campus- Wearable Mobile Devices

A Korean team designed and implemented a wearable computer, a mobile device in disguise of a cloth, from the scratch and named it UFC (Ubiquitous Fashionable Computer). In addition to that they developed a wireless gesture recognition device called i-Throw, whose device is comparable to a ring.[9] The Basic components of U-TOPIA are, besides the already mentioned wearable user device and the i-Throw technology, a java based middleware and by the wireless mesh network, that consists of a area-wide WLAN and network of ZigBee sensors. LBS are provided by the



(a) U-TOPIA: Wearable Devices (b) i-Throw: Receiving Data Gesture

Figure 2: U-TOPIA: Campus-wide Advanced ubiquitous Computing Environment [9]

WLAN, and for indoor tracking and low-power short message transfer ZigBee was adapted. The middleware was a product by another middleware team and it works among others as a location server, that gathers and manages the location information of both UFC users and ubiquitous devices, and it is also responsible for permission handling.

Whenever a user wears the jacket on which UFS is installed (see Fig. 2(a)), he is able to have control over various devices and he is also able to execute various operations on his own. For instance every time a user points to a device, the UFC platform displays the selected target device upon its screen, afterwards he can start using the device. Another use case is that he can exchange data with other users or data devices in the room by moving his arm and hands upon different predetermined patterns. (see Fig.2(b)) In this way he can leave messages or upload pictures to public displays. On the campus of Korea Advanced Institute of Science and Technology, where U-TOPIA was invented, one can buy electronic papers after pointing at a device called u-Kiosk and throw it away after reading by pointing at a device that has the function of a trash. Tasks like print files can be resolved easily. While one usually has to connect to a network printer with his notebook or smartphone, a UFS user just need to point at a u-Printer after selecting the file he wants to print, which is stored on his UFC .

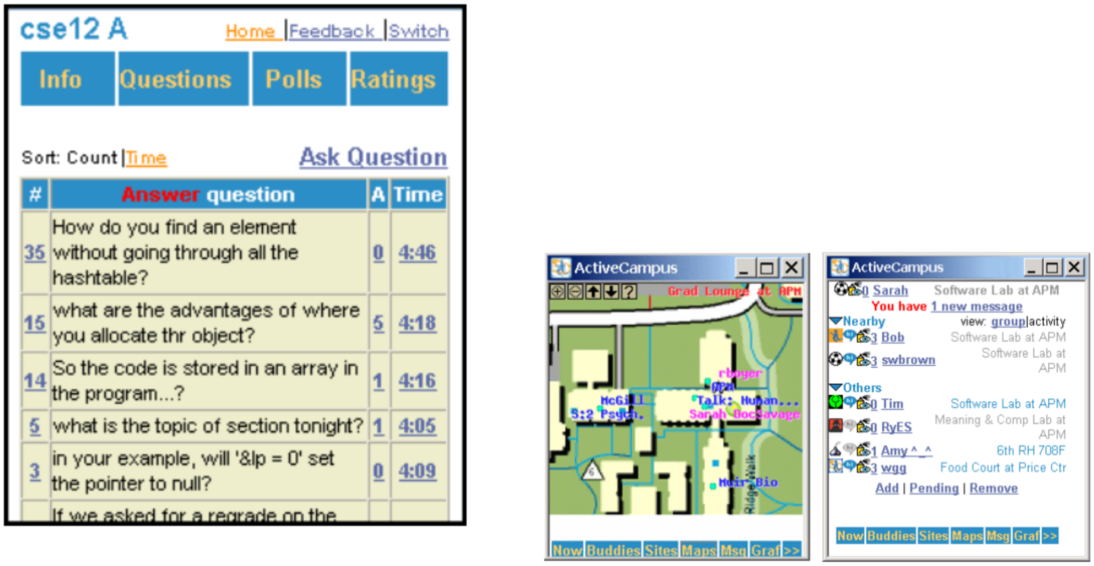
The Utopia team believes that U-TOPIA can be a realistic role model to improve current paradigm of ubiquitous computing environment within a few years.[9]

2.3.2 Real World But Virtual Identity - ActiveClass

Freshman students tend to keep questions during a lecture to themselves. Many prefer to leave the lecture hall without having understood without the essential, and most of them are too shy to raise the hand in order to pose the question, which probably is occupying many other students in the lecture hall too.

In order to resolve this well-known problem among freshmen, the idea was suggested, if the lecture would become better, if students can propose questions anonymously with the aid of a web application. The only equipment students need is a smartphone or a laptop with WiFi and a web-browser. The questions may be either anonymous or known of identity. In real time the question is sent and shown on the display, which is installed on the speaker's desk.

A project was launched in the year 2004 on the UC San Diego Campus in order to encourage classroom activities such as anonymous asking of questions, polling and student feedback. (see Fig. 3(a)) [8]



(a) Question list sorted by question vote count (b) Campus maps with buddies

Figure 3: ActiveCampus - Experiments in Community-Oriented ubiquitous Computing[8]

2.3.3 E-Social Network On Campus - ActiveCampus Explorer

As the title is self-explanatory this application refers to the idea, that students can add each other to their own buddy list and they can see anytime where their buddies are located and even to track them, provided that their buddy has allowed other persons to see his location.

When UC San Diego introduced ActiveClass in the year 2004, they also introduced a project called Active-Campus Explorer simultaneously. If a professor has to attend a meeting in 5 Minutes and he knows that everybody is in house exception one person. Coincidentally the missing person is on his buddy list, so the professor can open the buddy campus explorer to see where the missing person is. If he is approaching the meeting room, then this means probably that he is on the way. (see Fig. 3(b)) At that time the participating developers were confronted with problems like limited battery life, bad wireless lan connectivities, which often results in corrupted location detection and HTML without Ajax technologies, which means no server push and bad real-time presentation of buddies on the map. Nowadays apart from the problem with short battery runtime, other problems do not exist any more.

It is really amazing that already in the year 2004, UC San Diego offered a kind of "indoor navigation", if one considers that ActiveCampus merely detected location through the PDA's report of currently sensed 802.11b access points. It is worthy of mention that today spatial coordinates of the student can be gathered using mobile phone's in built global positioning system (GPS), RFID, WLAN Detection or some other unfamiliar techniques. GPS is used generally in the open air, and for indoor navigation it is common to use a hybrid building navigation system, consisting of several information booths and techniques. Having the student's position a smart search algorithm should be able to bring everybody through the campus to anywhere.

2.3.4 Public/Situated displays

Public displays can be used multisided. One can just simply uses it as a surface area for presentation, for example where pictures are shared, or he can use them as a notebook. (see Fig. 4) Lancaster University's e-Campus project is exploring the creation of large-scale networked displays as part of public, interactive pervasive computing environment. On the campus of Lancaster University, one could leave notes and messages to the office door displays, whenever the person he intended to visit was not in. In order to ensure that displays do not fall victim to vandalism, the display is often installed within the office whenever it is used as a notice board. The Lancaster University came to the conclusion,

Figure 4: Visitor using his mobile phone to interact with a public display using Bluetooth[11]

that the deployment of mass public displays is quite difficult. The difficulties arise from costly deployments, strenuous changes and maintenance. [12]

2.3.5 Canteen

Frequently it happens to a student, that he goes to the canteen and find a very long waiting queue. Regardless of whether it is dequeuing or enqueuing, standing in a queue is waste of time. By means of location based services, it can be detected how many smartphones (people) are near the canteen. Based on the scanned result, the service estimates the number of persons in the canteen, and then if needed, it sends a "warning" notification to students, who are approaching the canteen and finally it saves the students' IDs in a list. After dequeuing the service sends a "green light" notification to students who are on the list and has not been in the canteen since the last "warning" notification.

On the one hand the canteen is neither underutilized nor overutilized, but always utilized optimally. On the other hand students save time during their lunch break and can invest more time in studying.

3 Risks And Criticism

3.1 Security Aspects

The ubiquitous computing era also brings some dangers, the potential invasion of privacy being one of the more important issues. Indeed we are tracked permanently, and who knows what is going to happen with these tracking data? What if they fall into one's hands, who have criminal tendencies? Will Orwell's 1984 or Bradbury's Fahrenheit 451 become reality one day, if ubiquitous computing keeps on progressing like now?

Ubiquitous campus applications poses new security issues that are mainly privacy, usability and hybrid scheme. Privacy is an issue, because users use numerous public devices. They can be either explicitly used, with the user using input devices on the public device, or implicitly used in applications in which user's personal devices interact with the public devices. Usability needs to be taken into account when users begin to use numerous devices, and conducting security procedure on each of them becomes a burden. Hybrid scheme would be needed when there are multiple methods for security, such as passwords, identification cards and biometrics for authentication [10] Often it is not easy to find the right balance between usability and security.

3.2 Criticisms By Environmentalist

By deployment of so many technical devices, we have also to consider facets on the part of the environmentalist . They fear that these devices spend too much energy on the one hand, and on the other hand they claim that mobile devices cause diseases because of their electromagnetic radiation.

3.3 Necessity Of Ubiquitous Computing

Ubiquitous computing must engage the periphery of our attention and not be the center of our attention, said Mark Weiser. It must be taken into account, that some devices and services may demand period of vocational adjustment. No matter how plentiful ubiquitous computing resources are, it means nothing from the perspective of a user if a user cannot access the resource easily via a well-organized user interface. User interface should be easy to learn, easy to use and user-friendly and environment-friendy. But it must also have a groundbreaking design in order to attract young people. As already mentioned in 3.1, sometimes the handling of a new device is so complex, that it takes contra-productive effects. As a result, the period of use is short compared to the training period.

Also we have to consider, that it always takes time for developing a product until it is stable. Sometimes the product is of no use just after it has reached it's maturity stage because newer and more efficient technologies have been invented, which is better for a long period of time.

4 Conclusion

The main objective of ubiquitous campus is to make students more active and interactive. In this paper, we have itemlized many successful projects, especially foreign universities have broken the ground and done a good job. The author of this paper believes that ubiquitous campus not only increases a student's efficiency, but also provides learning content through seamless connectivity. It is a enrichment both for universities and for students.

References

[1] Weiser, M.: The computer for the 21st century. (1995) 2

[2] Weiser, M.: The future of ubiquitous computing on campus (1998) 3

[3] Niemelä, Eila and Latvakoski, Juhani: Survey of requirements and solutions for ubiquitous software. (2004) 2.1

[4] Hassan, Z.S.: Ubiquitous computing and android. Digital Information Management, 2008. ICDIM 2008. Third International Conference on (2009) 2

[5] Griffin R., Pyo G., Serotkin: Turning Vision and Expectation Into Reality: Building a Wireless Campus and Library for the Future at Saint Francis University (2001) 2, 2.2.2

[6] Nelaturu, S.C.B. and Kambham, R. and Karna, N.J. and Parupalli, R. and Mandula, K.: Building intelligent campus environment utilizing ubiquitous learning (2010)

[7] Hada, Y. and Shinohara, M. and Shimizu, Y.: K-tai Campus: university-sharing campus information system applicable to mobile phone and PC (2005) 2.2.1

[8] Griswold, W.G. and Shanahan, P. and Brown, S.W. and Boyer, R. and Ratto, M. and Shapiro, R.B. and Truong, T.M.: ActiveCampus: experiments in community-oriented ubiquitous computing (2004) 2.3.2, 3

[9] Park, K.H. and Lee, J. and Yoo, J.W. and Lim, S.H. and Park, K.W. and Choi, H.J. and Wohn, K.: U-TOPIA: Campus-wide Advanced Ubiquitous Computing Environment (2007) 2.3.1, 2

[10] Matsumiya, K. and Tamaru, S. and Suzuki, G. and Nakazawa, J. and Takashio, K. and Tokuda, H.: Improving security for ubiquitous campus applications (2004) 3.1

[11] Alan, K.C. and Dix, A. and Fitton, D. and Kray, C. and Rouncefield, M. and Saslis-lagoudakis, G. and Sheridan, J.G.: Exploring mobile phone interaction with situated displays (2005) 4

[12] Storz, O. and Friday, A. and Davies, N. and Finney, J. and Sas, C. and Sheridan, J.G.: Public ubiquitous computing systems: Lessons from the e-campus display deployments (2006) 2.3.4

[13] Jeon C., Lee M., Oh S.: A Design and Implementation of Ubiquitous Campus Environment Information Service (2008)

Interaction with Mobile Navigation Systems

Mathieu Davaine
mathieu.davaine@mytum.de

February 10, 2011

Abstract

This paper offers an overview of the interactions in existing mobile navigation systems. It presents a wide variety of systems classified by their output type: visual, haptic, audio and others. Each system is reviewed and compared to the others, exposing its features, field of applications and the standard requirements for navigation it meets. It will be seen that audio and haptic methods are able to provide a real alternative to the standard map visualization (e.g. when the visual sense is occupied by another task or when the user is visually impaired) and that basic concepts such as the usage of landmarks for pedestrian tend to improve the navigation task.

1 Introduction

Mobile navigation is no longer a dream since 2000 when the GPS was made available for precise civilian usage, allowing accurate outdoor navigation and location. A lot of mobile devices provide today's users with services such as navigational assistance and therefore have to face a lot of constraints [1]:

- Technical resources such as computational power, display capability.
- Dynamic environment (e.g. user constantly moving at different speed).
- User current cognitive abilities, if a user suffers for a cognitive impairment or needs a certain amount of situation awareness.
- User needs : is the user a tourist who wants a lot of information about his surrounding, or does he just want to travel to his destination as fast as possible ?
- Access to a precise user location: outdoor positioning can be obtained more or less precisely with Global Navigation Satellite Systems (GNSS), but those systems do not work well inside buildings. Moreover, user's current orientation is not directly given. That's why pedestrian navigation is also a lot more difficult than car navigation: pedestrian are more likely to be in urban corridor and the system can't use the road for the positioning as it will do with a car.

Today's solutions cannot fully cover all of these demands and have to make compromises to offer the user the best experience. Interaction between the systems and the user revolves around 2 components: user input to the system and system output to the user. Input is mainly characterized by route planning and user current position, speed (through GPS or similar) and orientation for most of the systems. Output is one of the major features of a navigation system because it is how it provides the user with feedback about his navigation choices. This is the feature that will be mostly covered by this paper as it exists in a lot of different forms, each of them suited for a particular situation. A single route instruction like 'Turn left into Main avenue' can be transcribed in many ways other than a text, like 2D/3D animations, speech, vibrations and even a combination of all three all in accordance with the current context.

Those different types of systems will be presented in the paper as follows. First the standard visual output with 2D and 3D maps but also more exotic ways involving augmented reality. Then haptic and audio feedbacks are presented. Finally, multimodal and others systems are shown followed by some concluding remarks.

2 Visual Feedback

2.1 2D Maps and Sketch

Before mobile navigation systems, maps where used since thousands of years and are still a powerful tool for representing an environment while navigating. Therefore maps on mobile device are still the primary output format when it comes down to navigation. However, the user has to cope with the small display size of the device and other restrictions (e.g. colors). This drawback is easily overcome by the interaction possibilities an electronic device can offer, such as marking the user position and highlighting specific roads or destinations.

In order to present the navigation information as clearly as possible, the map display should meet some requirements. Studies have shown that a simple digitalization of a paper is not the most efficient way to go: people prefer a low fidelity representation laying the emphasis on important objects such as the current user position and its surrounding [1]. Level of details can also evolve dynamically depending on the user's need: in the project REAL [2], the longer a user stays in an indoor area, the more information he gets about near objects (e.g. in a museum when you enter a room). Movement speed is also taken into account, the scale of the map being adapted to user's current state (running/walking).

Map orientation is also a factor in the navigation task. While a north-up map (north is always at the top) offers a better situation awareness (SA), a Heading-up map (automatically follows the user's orientation) is better suited for navigation [3]. It comes from the mental rotation needed in the case where the map is north-oriented. However the navigation systems need to know the correct orientation of the user, which might be hard to get as a pedestrian compared to a car moving a lot faster. This can be overcome by using an electronic compass but then again the constant rotation of the map might confuse the user and manually rotating provides better results [4].

To maximize the navigation efficiency, the system must know the current resolution of the user's state in order to adapt the display accordingly. The less information you have about user's position and orientation, the more information you need to display. In the indoor navigation system presented by Butz et *al*. [5], the graphical way description goes from a simple arrow when the user's position and orientation are well known to a wide map display with notable landmarks when the user's state is unknown. The user can also indicate his position himself to narrow down the representation. The need to display a large map without knowing the user's position can be a hurdle due to the small display size. Instead of displaying a small portion of the map and using the device's joystick to navigate it has been shown that a dynamic peephole navigation offers better results[6], where instead of moving a virtual map under the device's screen, the user physically moves the device over a virtual map. However more sensors in the device are needed to perform this task, or the usage of a virtual grid [7].

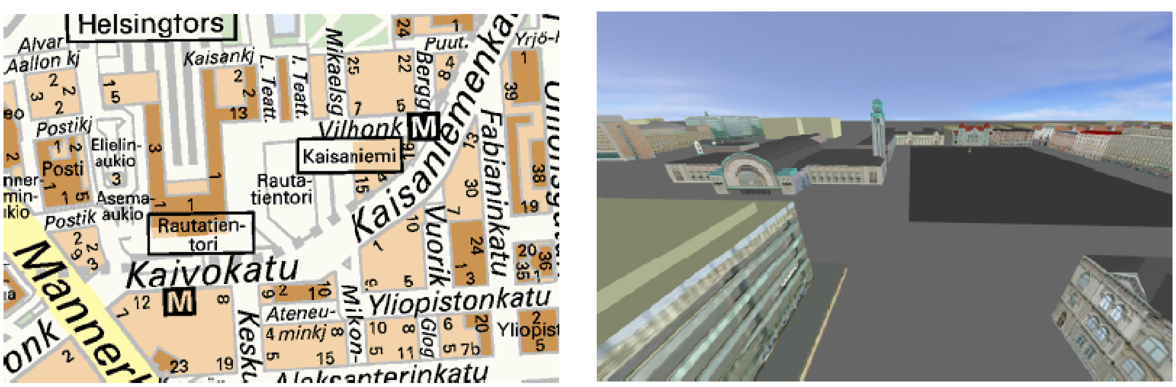


Figure 1: a 2D and a 3D representation of the same area. [8]

2.2 3D Maps

For pedestrian navigation, navigating with landmark offers better results than distance-direction instructions [9]. A 3D representation provides the user with a more realistic representation of his surrounding and thus helps him to identify noticeable landmarks on the output device. The major drawbacks of 3D systems are usually the small display screen of the devices and the technical workload needed to fully render a tridimensional environment. Manual Navigation presents also difficulties [8]: because of the added dimension, the usual arrows on a mobile device is not well-suited for this task. Thus 3D maps are better designed for an egocentric view (user centered) rather than an allocentric one (global representation of the route). This feature requires a precise knowledge of the user position and orientation which is sadly not always available in a pedestrian context. This is why 2D maps still tends to offer better results than 3D representation [10]. However, the 3D display can be seen as an addition to the 2D maps in order to provide the user with more information. However in driving condition where position and orientation are more accurate, 3D representations offer acceptable result. 2.5D map (tilted 2D maps) are often used to reduce the technical workload.

2.3 Augmented Reality

Augmented reality systems use footages taken by the user and add information for navigation. We can distinguish 2 types of systems: realtime (e.g. video) and delayed (e.g. the user takes a picture then wait for the image to be processed). This is the more resource demanding form of output as it has to perform a computer vision task on the environment, which takes a lot of computational power in the case of a video and this is why there are not many realtime systems as of today in small mobile units. The system presented by Gerhard Reitmayr and Dieter Schmalstieg [11] use an external laptop and displays the enhanced world through a the lens of an helmet. The user can see directly the waypoints blend into his environment, thus removing the need of changing his gaze by looking at a device. However, the system fails at his task when the GPS error becomes too high. Benjamin Walther-Franks and Rainer Malaka used prerecorded photographs instead of a videosteam in their system [12]. When a user crosses a waypoint, a photo of his current view is displayed with navigational information. The result of the study showed that this kind of navigation performs better as standard 2D maps in term of usability. However, errors occurred during the test, again due to a poor GPS signal.

Delayed systems benefit of having a lot of time to process data in order to provide a more accurate route navigation. Harlan et *al* [9] developed also a system based on photographs of noticeable landmarks, the difference being that in this case the user takes the photo. The image is then matched with other pictures of the same landmark which then are used to compute precisely the user position Fig.2. While this method is restrained to noticeable landmark, other systems follow the same pattern by using specific navigation signs like QR codes [13]. They act as waypoint were the user get navigation instruction, just like a system based on GPS without the positioning error but with less flexibility.

Figure 2: an ehanced picture taken by a user [9]

Another category of augmented reality devices is the one which enhances preexisting navigation systems. PhotoMap [14] for instance allow the user to take away a public map, and thus offers an up-to-date and precise description of his environment. Most of the general maps do not have enough details to allow navigation in certain areas or facilities, like a medical complex, a park or a zoo. With the assumption that the public map of the area has a realistic scale and a correct 'You Are Here' placement, the user can use it freely after a short implementation process in the mobile map.

Last but not least, this decade has seen the advent of Google Street View, Bing Maps Streetside and similar systems. They offer a 360 degree 'bubble' view for locations usually spaced out by 10 meters. While it looks like one of the best ways of presenting enhanced pedestrian information, these systems suffer from a few problems such as streets not being photographed/having outdated pictures, or the photos having a large size, hindering devices with small internet connection. Another problem is the difficulty to navigate manually from one bubble to another, e.g. when the positioning signal is too low to be used. However this last point has been overcome by Street Slide [15], which allow the user to zoom out of the bubble in order to have a complete panorama of the street and its point of interest by dynamically combining different views Fig.3.

Figure 3: StreetSlide: panoramic view of a whole street and its points of interest. [15]

3 Haptic Feedback

Looking at a display screen is not well suited while riding a bike. Taking a call on a mobile phone with a navigating application can also be troublesome because the user needs the audio output and can't look at the screen for navigation information. Haptic feedback offers a reliable solution while navigating when vision and audition are required for other tasks. Some systems use haptic vestments like suit, gloves or belt to provide the user with directional information, other system use the simple vibration of the device.

Haptic suits [3] and belts [16][17] provide the user with directional information through vibrations toward his next destination. While this system is completely 'hand and gaze free', it needs a precise knowledge of the body orientation and position. A small error can leads to a higher disorientation feeling than with other methods, that is why a multimodal approach is recommended with those systems [3][18]: the user then keeps the benefit of a reduced cognitive workload while looking at a 2D map helps him to understand his current route. Another advantage of these systems is the ability to guide a person in an environment without visual landmarks. For instance, most of the tests were realized on a field [17], which seems not a practical use at first sight, but in case of an impaired vision (blind people, night time) this feature really begins to shine.

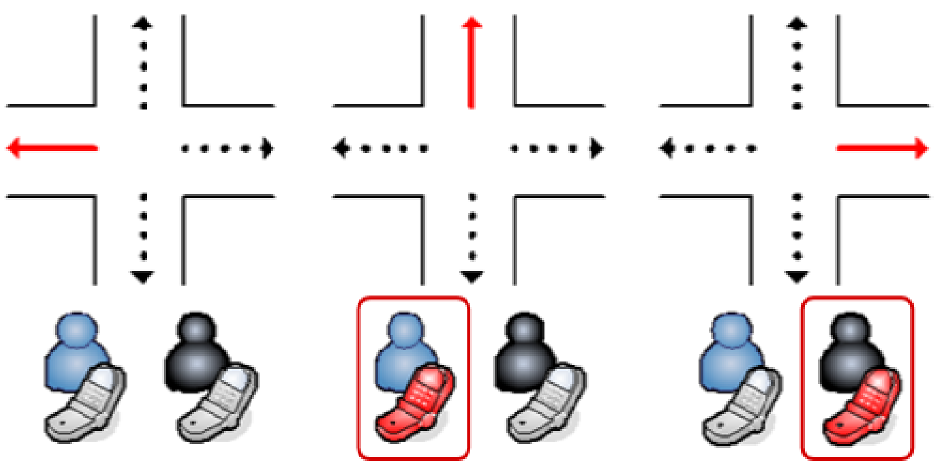


Figure 4: The Rotating Compass: The device of the user vibrates if the next direction of his route is lighted on. [19]

Route information can also be given to the user without the need of a specific device. Tactons (tactile icons) are a way to present information through a haptic rhythm. Navigation information can then be summarized with 3 different rhythms: left, right, stop and the distance to a waypoint can be encoded in the tempo [20]. The results show similar performance as a classic distance and turn audio feedback, but without the audio hindrance. Another way of navigating with simple vibration is demonstrated by Enrico Rukzio et *al* with their Rotating Compass [19]. The system presents two components: a vibrating device such as a mobile phone and a public display on every crossroads composed by blinking arrows, one for each direction. When the arrow corresponding to the correct route is lighted on, the user device vibrates Fig.4. The results of the study show that this system outperforms map and audio navigation in term of usability but at the cost of specific infrastructures.

4 Audio Feedback

As for haptic feedback, audio feedback can provide the user with an alternative source of information about navigation instruction. The most common audio-based navigation system is the speech-synthesizer which converts basic direction information into a human audible message and which can be found in pretty much all of the car navigation systems. However other systems exist for pedestrian usage based on music modulation, or 'speaking landmarks', which add a lot more context than sequential spoken instruction [21].

ONTRACK [22] and GpsTunes [23] are two systems using music modulation to provide the user with navigation information. Both need an additional orientation device such as a compass. The volume of the music heard by the user depends on the direction he is facing. The volume also increases when the user approaches the waypoint. This solution offers a pleasant and rather small cognitive demanding way of navigating, similar to a haptic device but without the need of a belt or vest.

Navigation through landmarks can also be possible as presented in the Roaring Navigator[24]. The user preselects a list of points of interest he would like to visit (here animals in a zoo Fig.5) and each landmark acts as a beacon with a specific sound. The limitation of this system is the permanent noise emitted by each landmark. However, it seems well suited for a group guiding application as the sounds are shared by all the users, thus preventing a social isolation that can be found in standard audio guide. Baus et *al* also explored the path of auditory landmarks produced by the environment itself [25]. The user is informed through any kind of feedback (e.g. text, image, speech) that an audio landmark (e.g. waterfall, cars, music) is present in its surrounding, thus adding more context to the navigation information. Most of the users found this addition helpful in their navigational task, but it cannot act as a substitute due to the lack of audio cue in certain areas.

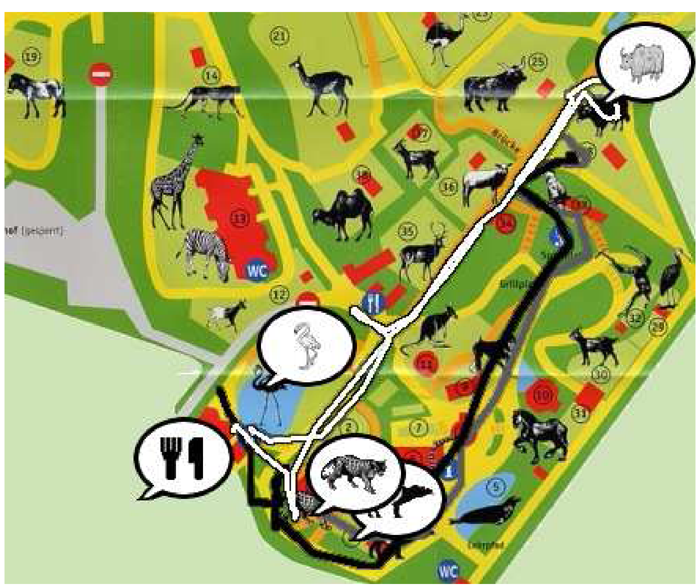


Figure 5: Map of the zoo where the auditory landmarks selected by the user are highlighted. [24]

5 Others

CityFlocks [26] is a social navigation system for visitors and new residents of an area. It consists in a web interface where user can share their experience about a location by offering a list of user comments. There is also a feature allowing the user to call or send a text message to a 'local expert' based on keywords. The navigational task is then performed by an experienced person rather than a machine. However, people seem resilient to engage a conversation with a complete stranger, thus limiting the benefit of this feature.

Other systems implement a multimodal approach: they combine all presented techniques in order to offer the user complete and customized navigational instructions. NAVITIME [27] is a pedestrian navigation system used a lot in Japan. It combines 2D, 3D and underground maps, voice and vibrating prompt. The European project HaptiMap [28] aims at providing users with a fully non-visual navigation system, visually impaired people being the first target group.

6 Conclusions

Features	**Text**	**Map**		**AR**	**Haptic**	**Audio**
		2D	3D			
Location	req.	depends	depends	req.	req.	req.
Orientation	req.	depends	depends	req.	req.	req.
Technical resources	very low	low	very high	moderate to very high	low	low
Information Quantity	low	high	high	high	low	low
Gobal awareness	inexistant	very high	very high	high	low	low
Cognitive resources	high	high	high	moderate	very low	low

Table 1: Comparison between the different methods.

Tab.1 presents the pros and cons of the reviewed methods. Most of the systems heavily rely on position and orientation instructions, which make them exposed to errors. Only a map representation allows the user to navigate manually. The technical workload, a crucial point for small devices with low computational power, is very high demanding in the case of video and 3D feedbacks. To be noted that haptic and audio feedbacks require sometimes additional devices (e.g. haptic belt, earbuds). Information quantity represents the 'Bandwidth' of the feedback: while a lot of information can be processed through vision, audio and haptic channel are less efficient, thus the navigation information must be kept to a strict minimum in those cases. Global awareness can be understood as the ability of the system to give the user general information about his surroundings. Here again, only maps can cover this demand as haptic and audio feedbacks only allow sequential navigational instructions. Cognitive resources or situation awareness is the cognitive workload of the systems. Visual feedback locks the gaze of the user on his navigating device and can even force him to stop. Haptic feedback outperforms here other methods due to the low usage of the haptic sense.

As none of the methods solve each problem, the best solution remains a combination of all forms of feedback with the ability to choose between various outputs. The next generation of mobile navigation systems will also benefit from the advancement made in positioning (e.g. upcoming GNSS Galileo), communication speed and processing capabilities, allowing faster and better navigational instructions.

References

[1] Baus, J., Cheverst, K., Kray, C.: A survey of map-based mobile guides. In Meng, L., Reichenbacher, T., Zipf, A., eds.: Map-based Mobile Services. Springer Berlin Heidelberg (2005) 193–209 10.1007/3-540-26982-7$_1$3. 1, 2.1

[2] Baus, J., Krüger, A., Wahlster, W.: A resource-adaptive mobile navigation system. In: IUI '02: Proceedings of the 7th international conference on Intelligent user interfaces, New York, NY, USA, ACM (2002) 15–22 2.1

[3] Smets, N.J.J.M., te Brake, G.M., Neerincx, M.A., Lindenberg, J.: Effects of mobile map orientation and tactile feedback on navigation speed and situation awareness. In: Proceedings of the 10th international conference on Human computer interaction with mobile devices and services. MobileHCI '08, New York, NY, USA, ACM (2008) 73–80 2.1, 3

[4] Seager, W., Fraser, D.S.: Comparing physical, automatic and manual map rotation for pedestrian navigation. In: Proceedings of the SIGCHI conference on Human factors in computing systems. CHI '07, New York, NY, USA, ACM (2007) 767–776 2.1

[5] Butz, A., Baus, J., Krüger, A., Lohse, M.: A hybrid indoor navigation system. In: IUI '01: Proceedings of the 6th international conference on Intelligent user interfaces, New York, NY, USA, ACM (2001) 25–32 2.1

[6] Mehra, S., Werkhoven, P., Worring, M.: Navigating on handheld displays: Dynamic versus static peephole navigation. ACM Trans. Comput.-Hum. Interact. **13** (December 2006) 448–457 2.1

[7] Rohs, M., Schöning, J., Raubal, M., Essl, G., Krüger, A.: Map navigation with mobile devices: virtual versus physical movement with and without visual context. In: Proceedings of the 9th international conference on Multimodal interfaces. ICMI '07, New York, NY, USA, ACM (2007) 146–153 2.1

[8] Nurminen, A., Oulasvirta, A.: Designing interactions for navigation in 3d mobile maps. In Meng, L., Zipf, A., Winter, S., eds.: Map-based Mobile Services. Lecture Notes in Geoinformation and Cartography. Springer Berlin Heidelberg (2008) 198–227 10.1007/978-3-540-37110-6$_1$0. 1, 2.2

[9] Hile, H., Grzeszczuk, R., Liu, A., Vedantham, R., J: Landmark-based pedestrian navigation with enhanced spatial reasoning. Pervasive (2009) 1–18 2.2, 2.3, 2

[10] Oulasvirta, A., Estlander, S., Nurminen, A.: Embodied interaction with a 3d versus 2d mobile map. Personal and Ubiquitous Computing **13** (2009) 303–320 10.1007/s00779-008-0209-0. 2.2

[11] Reitmayr, G., Schmalstieg, D.: Collaborative augmented reality for outdoor navigation and information browsing. In: Proceedings of the Second Symposium on Location Based Services and TeleCartography. TU Wien (2004) 53–62 2.3

[12] Walther-Franks, B., Malaka, R.: Evaluation of an augmented photograph-based pedestrian navigation system. In Butz, A., Fisher, B., Krüger, A., Olivier, P., Christie, M., eds.: Smart Graphics. Volume 5166 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2008) 94–105 10.1007/978-3-540-85412-8$_9$. 2.3

[13] Chang, Y., Tsai, S., Wang, T.: Wayfinding systems based on geo-coded qr codes and social computing for individuals with cognitive impairments. of Advances in Information Sciences and (2007) 1–7 2.3

[14] Schöning, J., Krüger, A., Cheverst, K., Rohs, M., Löchtefeld, M., Taher, F.: Photomap: using spontaneously taken images of public maps for pedestrian navigation tasks on mobile devices. In: MobileHCI '09: Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services, New York, NY, USA, ACM (2009) 1–10 2.3

[15] Kopf, J., Chen, B., Szeliski, R., Cohen, M.: Street slide: Browsing street level imagery. ACM Transactions on Graphics (Proceedings of SIGGRAPH 2010) **29**(4) (2010) 96:1 – 96:8 2.3, 3

[16] Tsukada, K., Yasumura, M.: Activebelt: Belt-type wearable tactile display for directional navigation. In Davies, N., Mynatt, E., Siio, I., eds.: UbiComp 2004: Ubiquitous Computing. Volume 3205 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2004) 384–399 10.1007/978-3-540-30119-6$_2$3. 3

[17] Heuten, W., Henze, N., Boll, S., Pielot, M.: Tactile wayfinder: a non-visual support system for wayfinding. In: Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges. NordiCHI '08, New York, NY, USA, ACM (2008) 172–181 3

[18] Pielot, M., Boll, S.: Tactile wayfinder: Comparison of tactile waypoint navigation with commercial pedestrian navigation systems. In: Pervasive Computing. Volume 6030 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2010) 76–93 10.1007/978-3-642-12654-3_5. 3

[19] Rukzio, E., Müller, M., Hardy, R.: Design, implementation and evaluation of a novel public display for pedestrian navigation: the rotating compass. In: Proceedings of the 27th international conference on Human factors in computing systems. CHI '09, New York, NY, USA, ACM (2009) 113–122 4, 3

[20] Lin, M.W., Cheng, Y.M., Yu, W.: Using tactons to provide navigation cues in pedestrian situations. In: Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges. NordiCHI '08, New York, NY, USA, ACM (2008) 507–510 3

[21] Kray, C., Elting, C., Laakso, K., Coors, V.: Presenting route instructions on mobile devices. In: Proceedings of the 8th international conference on Intelligent user interfaces. IUI '03, New York, NY, USA, ACM (2003) 117–124 4

[22] Jones, M., Jones, S., Bradley, G., Warren, N., Bainbridge, D., Holmes, G.: Ontrack: Dynamically adapting music playback to support navigation. Personal Ubiquitous Comput. **12** (October 2008) 513–525 4

[23] Strachan, S., Eslambolchilar, P., Murray-Smith, R., Hughes, S., O'Modhrain, S.: Gpstunes: controlling navigation via audio feedback. In: Proceedings of the 7th international conference on Human computer interaction with mobile devices & services. MobileHCI '05, New York, NY, USA, ACM (2005) 275–278 4

[24] Stahl, C.: The roaring navigator: a group guide for the zoo with shared auditory landmark display. In: Proceedings of the 9th international conference on Human computer interaction with mobile devices and services. MobileHCI '07, New York, NY, USA, ACM (2007) 383–386 4, 5

[25] Baus, J., Wasinger, R., Aslan, I., Krüger, A., Maier, A., Schwartz, T.: Auditory perceptible landmarks in mobile navigation. In: Proceedings of the 12th international conference on Intelligent user interfaces. IUI '07, New York, NY, USA, ACM (2007) 302–304 4

[26] Bilandzic, M., Foth, M., De Luca, A.: Cityflocks: designing social navigation for urban mobile information systems. In: Proceedings of the 7th ACM conference on Designing interactive systems. DIS '08, New York, NY, USA, ACM (2008) 174–183 5

[27] Arikawa, M., Konomi, S., Ohnishi, K.: Navitime: Supporting pedestrian navigation in the real world. IEEE Pervasive Computing **6** (July 2007) 21–29 5

[28] Magnusson, C., Tollmar, K., Brewster, S., Sarjakoski, T., Sarjakoski, T., Roselier, S.: Exploring future challenges for haptic, audio and visual interfaces for mobile maps and location based services. In: Proceedings of the 2nd International Workshop on Location and the Web. LOCWEB '09, New York, NY, USA, ACM (2009) 8:1–8:4 5

Mobile Interaction with the Physical World

Florian Buchmeier
flobuchmeier@mytum.de

February 10, 2011

Abstract

This paper gives an overview of the interaction methods with the real world when using a mobile phone - *pointing*, *scanning* and *touching*. In addition to that, several use cases are described, from the *Smart Environment* in your living room to *information heavy applications* like visiting a museum or taking a guided sightseeing tour.

After that, one of the three interaction methods, *touching*, is chosen and it is described how it can be realized with the use of *Radio Frequency Identification* or in particular *Near Field Communication*. Then you can read more about the existing uses of NFC in public as well as about current scientific projects which use NFC technology like *Touch & Connect*[1] and *Touch & Interact*[2]. At the end of this paper you will find a summary of our work and how important physical interaction with your mobile phone will be in the future according to the author of this paper.

1 Introduction

Mobile phones have become a common sight in our daily life and serve us in many ways. As of now, most of us use them as a means to communicate, for example when writing short messages or emails. Since the beginning of 2007, when the first iPhone was introduced by Apple, smart phones have become more and more popular. In fact, according to a study by Vision Mobile[3], 20 percent of all new shipped phones in Q1 and Q2 of 2010 were already smart phones.

Modern smart phones are not only used for communication with other people, but also for accessing or controlling other electronic devices. Currently it is for example possible to remote control your personal computer with any Symbian, iOS or Android[1] powered mobile phone using a Bluetooth or WLAN connection. This method of accessing data and controlling another device with your mobile phone is called *User Mediated Input*, an optional part of *scanning*. The other methods described in this paper are *pointing* and *touching*. The exact definition of those three terms and their usage will be explained in the following chapters.

Apart from that, the question arises which interaction method users prefer in a given context. You can read more about this particular topic in chapter 3, *Comparison of different Usage Scenarios*. But before that, the reader is given an overview of the three different physical selection methods, then presented two usage scenarios and in the end give a short summary of how *touching* is realized using Radio Frequency Identification and in particular Near Field Communication.

2 Physical Mobile Interaction Techniques

As mentioned before, there are several ways to interact with your environment using a modern mobile phone. Currently most of the interaction is restricted to *scanning*, e.g. when setting up a new Bluetooth connection with another mobile phone or when browsing for content on the Internet, since especially in Europe, there are no commercially successful techniques for the end user that allow him to interact with his environment using a mobile terminal like his cell phone.

When interacting with electronic devices in the near vicinity of the user, *scanning* can be quite time consuming. This is where more intuitive methods like *touching* and *pointing* provide a much more enjoyable user experience, but require a significantly larger set of data, like the direction the cell phone is pointing in. In the following, you will find a short description of each interaction method and how it can be realized.

[1] Mobile phone operating systems by: Nokia, Apple and Google

2.1 Pointing

Pointing is generally seen as a comfortable and intuitive way to address other objects. As the name implies, *pointing* is an interaction technique where the user aims at a smart object to select it. It is seen as natural and intuitive (easy to learn) since we are used to point at objects of desire in our daily life. The technical realization can be archived using several technologies, for example light beams like laser pointers or infrared beams in conjunction with photo diodes, as well as visual markers like Bar- or QR Codes[2] which require a significant amount of image recognition technology to work. Especially Barcodes and even QR Codes have become quite common in our daily life, since every consumer product has a Barcode on his label and most mail or delivery services use 2D codes (like QR Code) to identify and track their shipping goods. QR Codes are often used for advertising[3] as they are easily readable for modern smart phones even from a far distance (as long as the picture is big enough).

The biggest disadvantages of *pointing* are, that it is not seen as error proof and secure as other interaction methods like *touching* and user mediated input, as well as the fact that it requires a direct line of sight to the smart object. (see Rukzio et *al*. [4])

2.2 Scanning

Scanning is currently the most common technique, because it is easy to implement and works across obstacles like walls. When a user scans his environment with his mobile phone, he will get a list of all available devices and services near his location. The user is then free to choose, depending on his desire, which device he wants to connect to. After the connection is established, *scanning* requires direct input from the user. This is achieved either through a hardware keyboard or the touchscreen of the smart phone. Applications like Gmote[4] on Google's Android platform allow the user to remotely control his computer while emulating a mouse and keyboard on the mobile device.

Scanning relies heavily on wireless networking technologies like Bluetooth or Wireless Local Area Network . Since most mobile phones shipped right now are equipped with at least a Bluetooth module, *scanning* is already widely used. However, it is seen as a long winded process and only used when necessary because of its high knowledge requirements, as the user has to for example know where to find the menus, the name of the device he wants to connect to, and in some cases think of a difficult pass phrase.

2.3 Touching

When we speak of *touching* in our daily life, we mean connecting two physical objects by bringing them as close as possible to one another, which normally involves physical contact. In this context, *touching* does not require any form of physical contact and 0 to 10 centimeters of distance between the two objects are sufficient for technologies like Radio Frequency Identification (RFID) and Near Field Communication (NFC) to work.

A typical process is as follows: In the first step, the user brings the two mobile devices close to each other. In the second step, the user is presented with a list of services this particular smart object or mobile device he just *touched* offers. He then has to choose one option from the list, like getting more information (e.g. an online user manual) or changing the state of the object (e.g. playing music from the smart phone on your home stereo by simply placing the phone on top of the home stereo).

This interaction method, like *pointing*, is seen as intuitive and natural, since we constantly touch things in our environment we want to interact with. However, touching is not always the easiest way to interact with a smart object, since depending on the situation it may require a high physical effort for the user to get close to the object.

2.4 Summery

As you can see, each interaction method can be realized using a variety of different technologies, and every one of them got his advantages and disadvantages. The next part will be exactly about two different usage scenarios and the users choice of interaction methods in a specified context.

[2] http://www.qrcode.com Retrieved on 2010-12-16

[3] http://blog.purevisibility.com/2010/03/mobile_marketing_and_how_to_make_the_barcode_hip_again/ Retrieved on 2010-12-16

[4] http://www.gmote.org Retrieved on 2010-12-16

3 Usage Scenarios

In this paper, two different usage scenarios where a user can interact with physical objects in his surroundings via *touching*, *pointing* or *scanning* are presented.

The first will be a *Smart Environment*, for example when the user is at home and wants to interact with his personal electronic devices. The second will be about two *Information Heavy Situations*, in particular when visiting a museum and taking a guided tour.

3.1 Smart Living Environment

A *Smart Environment* is, according to Mark Weiser "a physical world that is richly and invisibly interwoven with sensors, actuators, displays, and computational elements, embedded seamlessly in the everyday objects of our lives, and connected through a continuous network". [5]

Most of the time, people are either at work or at home, where a lot of objects want to be controlled or influenced in any way. For example, you may want to turn on the water for the bath tube while still relaxing on the couch in the living room, or get an online manual for a new electronic gadget you have just bought but do not want to use a computer for this task. All of this can be achieved using a smart phone and smart objects. The question is, which selection method is the most comfortable to use in the given situation.

According to a study by Rukzio et *al*. [6] where the participants had to select and switch on a CD player and after that, the heating, as well as access two links that were stored on a radio and a laptop, showed that the users choice is influenced by a variety of factors. At first, it depends depends heavily on the distance between the user and the smart object. When the user is close to the object, he prefers *touching* because of its ease of use and the low cognitive effort. When the distance increases and the user has to change his state of activity, e.g. from lying to standing and walking towards the object, he prefers *pointing*, as long as there is a line of sight from the user to the desired smart object. When the smart object is located in another room or there is no line of sight, almost all users switch to *scanning*, which requires a relatively high cognitive effort, but very low physical action.

3.2 Information Heavy Situations

Information heavy situations like taking part in a guided tourist tour or visiting a museum can benefit greatly from the use of modern smart phones to interact with the users surroundings.

Rukzio et *al*. [4] showed, that when walking through a park on a guided tour, user mediated object interaction was the simplest and at the same time the least enjoyable one, whilst *pointing* and *scanning* were the most fun according to the participants of the experiment (Touching was not an option in this test).

Things changed when the participants were sent on a guided museum tour using their RIFD enabled smart phones. Here pointing and touching are seen as innovative and enjoyable and user-mediated object interaction as well as touching were the most reliable interaction methods. In the end, almost everyone of the participants prefered touching as their interaction method of choice. This may be due to th fact that when being in a museum the user is already close to the smart object he or she wants to interact with, so the biggest disadvantage of touching, the high physical effort, has no impact in this context.

In addition to that, a study by Hsi et *el*. [7] showed, that the use of RFID technology can improve the visitors learning experience. This is achieved by giving the visitors RIFD cards that they can use to bookmark exhibits or let souvenir photographs be taken from nearby cameras. When they are at the kiosk or at home, they can access a dynamically generated web page with all the photos and information about the exhibits they visited as well as get additional information.

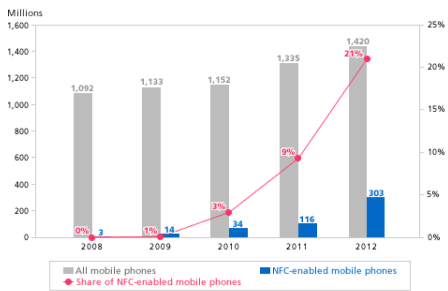
This feature can also be useful for teachers taking students on a museum field trip as it helps them keeping track of what material they have covered and tells them where to find additional information about the exhibits if needed.

Examples for RFID enhanced museums are the Exploratorium, a hands on science museum in San Francisco, the "Net World" exhibition in the Museum of Science and Industry in Chicago as well as the Museum of Natural History in Aarhus, Denmark, where stuffed birds contain RFID tags that reveal more information and media about the respective species.

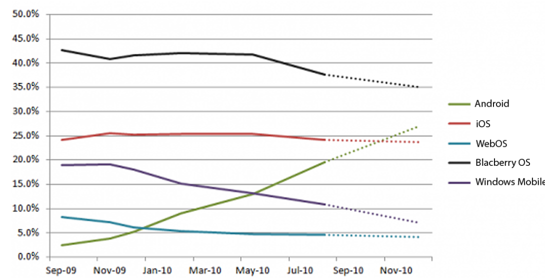
4 Touching realized through NFC

After this small overview of the different methods of interaction and their use cases, one method is discussed in detail. As mentioned before, *touching* is mostly realized using the Radio Frequency Identification (referred to as RFID) or Near Field Communication (NFC) technologies. RFID and NFC enabled mobile phones have lead a relatively shadowy existence up till now, with only a handful of supported devices available worldwide. But according to analysts at ABI Research (see Fig. 1(a)) and experts at Nippon Telegraph and Telephone DoCoMo (NTT DoCoMo), the number of NFC-equipped mobile phones will raise to 21% of all mobile phones in 2012.

Apart from that, Google's latest mobile operating system, Android 2.3 Gingerbread, officially introduces NFC support into the smart phone market. And since Google's market shares are raising since its first debut on the smart phone market in October 2008 (see Fig. 1(b)) another driving force will surely accelerate the introduction of NFC technology to the mass-market. In the next part, chapter 4.1, a brief description of RFID and its history is given. Then you can read more about the emergence of NFC and its applications as well as two scientific studies, namely *Touch & Connect*[1] and *Touch & Interact*[2].



(a) Red: Market share of NFC-enabled mobile phones according to mobile operator NTT Docomo[8]



(b) Green: Androids' market share rising from 0% to 20% in only two years since its introduction in October 2008. [9]

Figure 1

4.1 RFID

Radio Frequency Identification, in short RFID, uses electromagnetic waves to exchange data usually between a powered RFID reader and a passive RFID tag. It is mostly used to track or identify people, animals and assets. In fact, there is no world wide standardization for RFID, partly because of the different frequency bands used and their regulations. There are several International Organization for Standardization (ISO) standards, from ISO 14223 which is used for the identification of animals to ISO/IEC 15693 which is widely used for non-contact smart payment and credit cards.[5]

An early and maybe the first work about RFID was published by Harry Stockman[10] in 1948, called "Communication by Means of Reflected Power". Stockman already foresaw that "considerable research and development work [had] to be done before the remaining basic problems in reflected-power communication are solved".

Over ten years later, the first commercial activities were beginning. In the late 1960s, interest grew in electronic article surveillance (EAS) equipment for anti theft protection. Those single-bit EAS tags were small and could be easily attached to any kind of articles. It then took another 20 years for the final breakthrough of RFID technology, when integrated CMOS chips and the emergence of the personal computer (PC) were inevitable.

As of now, RFID is used in a variety of fields, for example in electronic toll systems (1991 in Oklahoma), to track goods or even for electronic payment with your mobile phone[6]. In addition to that, the market for RFID is constantly growing. In 2007, China alone spent $1.9 billion on RFID products, mainly for contactless national identification cards.[7]

[5] http://en.wikipedia.org/wiki/Rfid#Regulation_and_standardization Retrieved on 201-12-19
[6] http://www.rfidjournal.com/article/view/7224 Retrieved on 2010-12-20
[7] http://www.rfidjournal.com/article/view/6808 Retrieved on 2010-12-20

But exactly how does RFID work? At foremost, every RFID system consists of a reader (see Fig. 2(a)), which is typically used in conjunction with a personal computer, and a tag (see Fig. 2(b)), which is attached to an object and can store a variety of data, for example the price of the object, its owner or its origin.

The process is as follows: The reader outputs radio-frequency signals which for one, provide as means of communication with the tag as well as the energy required for the tag to communicate with the reader (in the case of passive RFID tags). The fact that passive tags do not require any batteries or power supply is one of the biggest advantages of the technology. Compared to optical read technologies like barcodes, the tag can be hidden and does not need to be on the surface of the object and therefore may be protected from wear. In addition to that, a single reader can read a large number of tags at once, instead of reading item by item. RFID tags are available in every imaginable shape and size, as small as 0.05mm x 0.05mm![8]

(a) A generic RFID Reader

(b) A RFID Tag used for electronic article surveillance

Figure 2: Typical RFID devices. Source: http://commons.wikimedia.org

4.2 From RFID to NFC

Near Field Communication technology was first developed by a joint venture of Philips (now NXP Semiconductors) and Sony in 2002. It basically combines the Radio Frequency Identification technology with smart card features, that allow the device to read and be read, as well as write and be written by another device. NFC was designed to enable intuitive communication between two devices and since the devices have to be really close together (a maximum of 10 centimeters) there is almost no risk of the data being eavesdropped. The NFC standard[9] was first issued in 2003 and is standardized in ISO 18092 as well as compatible to ISO/IEC 14443 (proximity cards used for identification), ISO/IEC 15693 (vicinity cards, can be read from a greater distance than proximity cards) and Sony's FeliCa, a smart card ticketing system (primarily used in Japan).

In contrast to RFID, where only one interaction method is allowed, respectively a reader to read or write a predefined tag, NFC allows three different mobile interactions, shown in Figure 3 below.

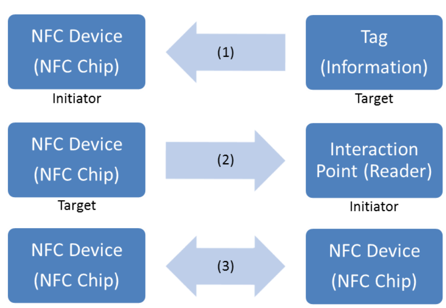


Figure 3: Different interaction methods for NFC enabled devices. In (1), the mobile device, initiates the data transfer by sending a RFID signal to the tag. In the next step, the tag responds and sends the information it contains back to the device. This interaction method is congruent with how a normal RFID process looks like. In (2), the NFC enabled device acts as the tag (or smart card) so that its information can be read by an interaction point. (3) shows that direct communication between two NFC devices is also an option.

[8] http://www.engadget.com/2007/02/14/hitachis-rfid-powder-freaks-us-the-heck-out Retrieved on 2010-12-20

[9] http://www.nfc-forum.org Retrieved on 2011-01-02

Some key features of NFC are, that it is compatible to existing infrastructures like Sony's Felica or exisiting RFID tags. Apart from that, its easy of use (the two devices only have to be close enough for the technology to work) and the inherent security, because of the short working distance, are very important as well.

4.3 Applications

Near Field Communication technology is currently tested in a variety of fields. The first large scale trial of the technology with 200 participants was launched ion October 18th 2005 in the city of Caen, France.[10] There it is used to simplify transactions in supermarkets (Monoprix) and fashion stores (Galerie Lafayette), for car parking tickets, for tourist information (landmarks and sights are equipped with RFID tags that contain for example a telephone number to get additional information about the sight) as well as active posters (e.g. for new movies or consumer products). All of this is achieved by using a prototype of Samsung's D500 mobile phone, which sports and embedded Philips NFC chip.

The first commercial adoption of NFC, namely get»in, was done in the Frankfurt Rhine-Main Region in Germany by its local transportation company RMV (Rhein-Main-Verkehrsverbund) in conjunction with Philips, Nokia and Vodafone,in April 2006.[11] The RMV in cooperation with mobile phone operator Vodafone currently only sells Nokia's 3220 NFC enabled cell phone at its stores in Hanau. The phones can be used for electronic bus tickets and in addition to that, they entitle the user to get discounts at local retail outlets and attractions.

Future applications include electronic payment and ticketing (see get»in), user friendly pairing of electronic devices (e.g. Bluetooth or WLAN key exchange), information gathering (e.g. reading product history) or asset management (reading smart tags with NFC phones for inventory control.

4.4 Studies

In 2008, Hardy et *al*. [2] studied the use of "Touch and Interact", a combination of a NFC enabled smart phone and a dynamic display, which consisted of a video projector and a specially designed silver screen. To sense the *touching* of the cell phone, this 50cm x 50cm silver screen was equipped with a mesh of 10cm x 10cm NFC tags and thus had an input space of 10x10 pixels. In this experiment, six different methods of interaction where implemented.

Hovering over a tag selects the underlying tag and the object on the virtual desktop. *Selection* works the same way, but requires the user to press a specified button. If the button is held, *Multi-Selection* is initiated. With *Polygon-Selection*, the users can now draw a polygon and select all enclosed tags at once. *Pick-and-drop* allows items to be picked up and dropped elsewhere, whereas *Remote Clear* is used to unselect any currently selected tags. *Pick-and-drop*. All of them include haptic and audio feedback for more assertive feedback during the selection of a tag. *Remote interaction* proved to be significantly worse in every way. The conducted user study then compared three methods of interacting with a display, namely *finger interaction*, like a conventional touch screen in combination with a mobile phone, *remote interaction*, where the phone acts as a remote control and the previously described methods of *Touch & Interact*.

The first part of the study showed that *Touch & Interact* ranked almost as good as *finger interaction* in terms of physical effort, average selection time and average error rate. Problems surfaced when the size of the object which had to be selected was too small (smaller than a single NFC tag) regarding *Touch & Interact* and when the distance between the user and the display grew larger, in the case of *remote interaction*. In the end, *Touch & Interact* was 135% faster than *remote interaction*. The second part focused manly on usability aspects and not outright performance. Because of the poor results of *remote interaction*, it was excluded from the test and only *finger interaction* and *Touch & Interact* where compared. Here, the users had to down- and upload files between the interactive display and their mobile phones as well as pick/drag and drop items on the screen. The study showed that eight out of twelve participants where in favor of *Touch & Interact*, because they found it easier to use, more intuitive and more fun to work with, although the users were not familiar with the technology.

In 2009, *Touch & Connect and Touch & Select*[1] by Seewoonauth et *al*. followed a similar approach and showed, that phone - computer interaction can be greatly improved by the use of NFC technology. In the corresponding user study, the developed method ranked significantly higher than a standard Bluetooth connection, because it was much easier and faster to use, especially for ingenuous users.

[10] http://www.nxp.com/news/content/file_1193.html Retrieved on 2011-01-05
[11] http://www.rmv.de/coremedia/generator/RMV/Tickets/Fahrkartenverkauf/getin/inhalt.html Retrieved on 2011-01-07

5 Conclusion

Simplicity and intuitive interaction with electronic devices has always been a key feature for their success on the market, since most users do not have the time or the will to study lengthy user manuals or learn how to use the device by try. This is especially true for information systems like computers and nowadays even cell phones, who evolved from simply means of communication to portable "super computers". Here, new, or rather "old" , methods of interaction with the real world must be found. In this paper, three different methods, *pointing*, *scanning* and *touching* are presented. It showed, that each interaction method has its own range of applications.

Pointing for example, is intuitive and preferred when the smart object is in line of sight and does require a change of state of the user, but is also more error prone than the other two methods. It is currently realized using Bar- or QR Codes and used for the tracking of goods, advertisement and of course product identification. *Touching* however, is seen as secure and error proof way to address and connect objects, as long as they are in touching range of the user. It can be achieved by using RFID or NFC technology and is actively used for inventory management, electronic payment and ticketing. *Scanning* is technically the most easiest way to interact with other objects in the near vicinity of the user, since it relies on existing wireless communication systems like WLAN or Bluetooth. For the end user, it is rather hard to use compared to the other two interaction types, since it requires a great deal of knowledge. On the other hand, it works across boundaries like walls or floors and does not require the user to change his position. As of now, it is used for connecting varies devices, like PCs or mobile phones. *Touching* crystallized as a particularly interestingly method and thus, its uses and realization were discussed further in the process of this paper.

Overall, a lot of work still needs to be done before any of the three interaction methods using smart phones will be accepted by the populace. Further information can be found in any of the given References below.

References

[1] Seewoonauth, K., Rukzio, E., Hardy, R., Holleis, P.: Touch & connect and touch & select: interacting with a computer by touching it with a mobile phone. In: Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services. MobileHCI ’09, New York, NY, USA, ACM (2009) 36:1–36:9 (document), 4, 4.4

[2] Hardy, R., Rukzio, E.: Touch & interact: touch-based interaction of mobile phones with displays. In: Proceedings of the 10th international conference on Human computer interaction with mobile devices and services. MobileHCI ’08, New York, NY, USA, ACM (2008) 245–254 (document), 4, 4.4

[3] Kapetanakis, M.: Smart < feature phones = the unbalanced equation (100 Million Club series). (2010 (accessed December 14, 2010)) http://www.visionmobile.com/blog/2010/10/smart-feature-phones-the-unbalanced-equation-100-million-club-series/. 1

[4] Rukzio, E., Broll, G., Leichtenstern, K., Schmidt, A.: Mobile interaction with the real world: An evaluation and comparison of physical mobile interaction techniques. In Schiele, B., Dey, A., Gellersen, H., de Ruyter, B., Tscheligi, M., Wichert, R., Aarts, E., Buchmann, A., eds.: Ambient Intelligence. Volume 4794 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2007) 1–18 10.1007/978-3-540-76652-0_1. 2.1, 3.2

[5] Cook, D.J., Das, S.K.: Smart environments: technologies, protocols, and applications. John Wiley and Sons (2005) 3.1

[6] Rukzio, E., Leichtenstern, K., Callaghan, V., Holleis, P., Schmidt, A., Chin, J.: An experimental comparison of physical mobile interaction techniques: Touching, pointing and scanning. In Dourish, P., Friday, A., eds.: UbiComp 2006: Ubiquitous Computing. Volume 4206 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2006) 87–104 10.1007/11853565_6. 3.1

[7] Hsi, S., Fait, H.: Rfid enhances visitors’ museum experience at the exploratorium. Commun. ACM **48** (September 2005) 60–65 3.2

[8] NTTDocomo: Waving at the Future with NFC. (2009 (accessed December 14, 2010)). http://www.nttdocomo.com/features/nfc/. 1(a)

[9] Vildosola, A.: The invasion continues, Android's market share skyrockets. (2010 (accessed December 14, 2010)) http://www.droiddog.com/android-blog/2010/10/the-invasion-continues-androids-market-share-skyrockets/. 1(b)

[10] Stockman, H.: Communication by means of reflected power. Proceedings of the IRE **36**(10) (1948) 1196 – 1204 4.1

A Survey of Interaction Techniques for Public Displays

Felix Greimel
F.Greimel@mytum.de

February 10, 2011

Abstract

This paper is concerned with different approaches for interaction possibilities with public displays. It mentions several methods about how people can retrieve information, share news or exchange data by using interactive displays. Privacy issues are emphasised as well. The way interaction is granted is distinguished by the provided input possibilities which are "interaction based on input devices", "interaction based on direct touch technology or haptics" and combinations of these two methods.

1 Introduction

Having pursued developing plasma displays, LCD displays and digital projectors, industry is now able to provide mature technology at reasonable prices. More and more displays appear in public places like lobbies, shopping malls, pedestrian areas, subways, airports, train stations and in many more. In contrast to CRT display technology it is now affordable to install large displays in confined space, due to the "flat screen", lower initial costs and less running costs because of smaller power consumption.

These displays usually show general information like time tables at train stations or cinema lobbies, possible activities in tourist information centers, news or advertisements in subway stations and so on. There are many application fields. The spectator normally stays passive because today's displays still are mostly not interactive. There are several approaches to involve the person looking at the screen. That way the user or even multiple users, if multiple input is provided, can interact and retrieve more detailed, more interesting or, if desired, personal information. But which method of user input is the most convenient, intuitive, practicable and yet realizable? In this paper three principles of input means are described. A distinction will be made between interaction using input devices, using gestures and combinations of these two methods.

If an identification method is included in the display system, it is even possible to present only information which is valuable to the user, like a personal schedule, if the display is placed in a large office [1], or personified news/advertisements concerning the users' interests. The question arises how privacy can be kept up. Different solutions to this issue will have to be found according to the individual approaches.

2 Interaction Based on Input Devices

Traditionally, interaction between people and a computer is handled by one or more input devices, e.g. hardware like keyboard and mouse, and a display giving optical feedback. With large public displays this kind of input might not be appropriate anymore. This is due to several reasons. One of those reasons is that it is impractical to have some kind of desk with mouse and keyboard right in front of a screen which is 50" wide or even more. This system similar to a personal computer would consume more space which annihilates one of a flat screen's advantages: saving space. Additionally, it is very likely that the user might need to take a step back in order to follow the cursor on a large display, dragging him away from the input device. Nevertheless it is still frequently seen that a system consisting of a projector and a notebook is used for letting people interact with large displays. H. Brignull and Y. Rogers used such an installation to examine on how to entice people to interact with large public displays [2]. They used special software and a common input possibility which made interacting easy for a single person. However, there are more inspiring options to increasingly involve people in an active way, like new exciting technology which provides more intuitive input for multiple users.

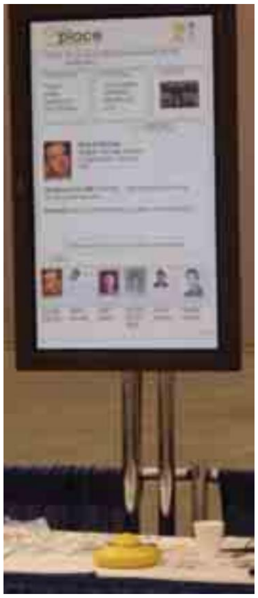


Figure 1: The PlasmaPlace display can be controlled by a large trackball.

A different approach was made by Churchill et *al*. They published a paper [3] in 2004 about a method to create awareness of online social activities in physical community places. At two different conferences (CHI and CSCW) they used a large-screen display in an upright position (Figure 1) to present the corresponding online community (CHIplace / CSCWplace). The display's mixed content, amongst others consisting of schedules, photos, blogs and people's profiles could be browsed through by a large trackball "to simplify the installation" [3]. Visitors could use the trackball's two buttons to change the display selection to the next or previous item in a given order. Spinning scrolled the selected display.

But a single trackball as in the mentioned scenario still limits the number of interacting people to one. This could be solved by providing multiple input devices managing individual cursors. Such an approach is described by Pawar et *al.* [4] in a project for pupils' interaction with a "digital black board". In this way multiple users are able to interact simultaneously with content being viewed on a large display, using cursors of different colours. In [5] Izadi et *al.* published a suitable user interface for public interaction with displays and supporting cooperative sharing and exchange of media. This UI provides also individual cursors.

At the second conference (CSCW 2002) Churchill et *al.* slightly improved their display's input technique by replacing the trackball by a touch sensitive overlay placed upon the screen surface. Thus, input was possible in a much more intuitive way. But still, parallel actions by two or more users were not realizable.

The presented scenarios are suitable for indoor environments like conference halls, shopping malls or offices. Concerning those locations vandalism or changing weather conditions usually aren't an issue, so input devices don't need to be as robust. But considering systems placed in an outdoor public area, like pedestrian areas or less monitored places e.g. outside tourist information centers at a national park, vandalism and changing weather conditions must be taken into account. This may be another reason for a keyboard and mouse combination being "outdated" or at least inappropriate. This means that the input devices of such interactive systems must be designed in a robust way to withstand the mentioned aspects as well as heavy use.

R. Ballagas et *al.* [6] published a possibility to use one's own mobile phone to interact with public displays which are inaccessible for direct touch based interaction, since they either need to be protected from vandalism or are simply out of reach. Today's mobile phones are usually equipped with a camera which is considered as a powerful input channel. Another standard feature of mobile phones is the wireless technology Bluetooth. Using the combination of these two features the presented interaction method *Point & Shoot* could be implemented (Figure 2). This method for selecting objects on the screen, is based on visual codes. Taking a picture of the initial code establishes a Bluetooth connection between the user's phone and the display. This is reported to result in a very low threshold for starting using this technology.

After the connection has been built up a matrix of codes is the base for further interaction. The immanent information of the codes establishes an absolute coordinate system on the display surface. The two dimensional codes can be read

Figure 2: Point & Shoot technique: (Left) The phone display is used to aim at a puzzle piece on a large display. (Middle) Pressing the joystick indicates *selection* and a visual code grid flashes on the large display to compute the target coordinates. (Right) The grid disappears and the targeted piece highlights to indicate successful selection.[6]

independently of the perspective distortion arising from the mobility of the phone camera. Until now, those codes are only displayed in the moment of selecting an object in order to minimize the distraction. It is mentioned, that future displays will be able to display codes in infra red, so that they can be read by the camera but are invisible to the user. Aiming at the screen by using the phone camera produces a live image on the phone which acts as a viewfinder. Thus, one can point at a desired object and "shoot" a picture by pressing the phone's joystick button. Now the decoded information of the code which was overlaid on the display in the moment of "shooting" will be sent to the display. Having processed this information, the computer connected to the display now knows the coordinates of the desired object and selects it.

R. Hardy and E. Rukzio presented 2008 an interaction technique combining an arbitrary screen equipped with a grid of NFC/RFID tags and a NFC (Near Field Communication) phone [7]. This technique enables the user to select contents of the screen, like textual information or photos, by touching the corresponding area with his mobile phone. When the phone touches the display the distance to one of the NFC/RFID tags is minimal, the phone sends a message like "tag xx was touched" via Bluetooth to the display server and the display gives visual feedback according to the coordinates or transfers the desired content wirelessly to the mobile device. Like the *select & pick* option which is used to download content to the mobile phone there is also a possibility to upload data. With *select & drop* media stored on the mobile phone can be selected on the device and dropped onto the display by touching it.
One main challenge of this technique is the relatively low input resolution. In this scenario the touchable cells measured 4cm x 4cm. However, this can be solved by using smaller RFID tags or, as Hardy and Rukzio suggested if multiple options are available within one cell, use either the phones joystick or keypad or an enlarged version of the selected area. In comparison to usual touch technology this technique offers one main advantage: the mobile phone can show additional, potentially private information on its screen. In this way sensitive information such as passwords or address information can be presented discretely. Besides that, the storage capabilities of mobile phones as input devices can be used for phone identification, supporting multiple users simultaneously and authenticating users regarding access privileges and billing.

A simpler approach is described by Cheverst et *al.* in their paper [8] about exploring mobile phone interaction with situated displays. They created a graphical user interface for displays situated in universities or offices that can be interacted with using again Bluetooth technology and a mobile phone. Users can easily connect to the public display by searching for surrounding Bluetooth devices and pairing with the corresponding display device. After a connection has been established it is possible to receive vCards or photos. Users can also upload pictures. This provides the possibility for users owning a phone equipped with a touch screen to "scribble" and leave a message by uploading the file of this note. In Cheverst's paper also a more synchronous interaction style is presented. Therefore the user can download an application called *vcfDownloader*. With this small app, the user can either select a person's vCard by browsing trough an array of user profiles using the joystick and the OK button or as a second approach, by using appropriate buttons for the ID representing the person of interest. Method number three consists of a graphical user interface on the mobile phone. Now one can select a picture relating to the vCard that the user wishes to have added to his contacts list. In a second paper by Cheverst et *al.*, which is about the possible positive contributes of a public display to a community, they use the same display system as a platform for publishing and sharing pictures of the concerned community, e.g. a sports club at a university. Sharing is done by up- or downloading pictures in a similar way as described above.

In Boring et *al.*'s paper [9] a technique is described, which combines many different displays using a mobile phone equipped with a touch screen and a camera. Contents on displays which are usually unreachable can be manipulated remotely. Screens that are incapable of touch interaction, such as wall projection or older screens can now be interacted with. Using a mobile phone camera one can aim at a screen and select an object in the live video by pressing the phones touch screen. Further on the user can transfer the selected content to a different screen by aiming at the desired one (Figure 3). This is realized by a server managing the displays and handling the pattern recognition by using the phones camera via a wireless LAN connection.

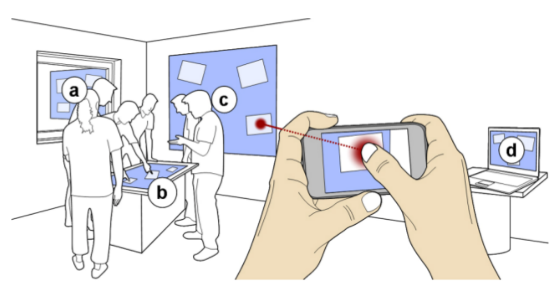


Figure 3: The TouchProjector [9]

Waldner et *al.* published a paper about an interaction technique which includes optical lenses in combination with polarising projectors. Two projectors can create two completely different images on one single screen which can be separated by using polarising glasses. This enables two groups of users to work on one big screen without interfering with each other. However, people who are not wearing the mentioned glasses are able to see a mixture of the two provided images. This results in a lack of privacy which should be considered when choosing the content to be displayed. Waldner et *al.* use a rear projection setup which avoids shadows of people standing in front of the screen. Additionally they provided a camera with a polarisation filter to get a clean picture of users' annotations at the screen for post processing. Another option is to collaborate and work on different layers. The "magic lens" makes it possible to view parts of the image provided by the second projector (Figure 4). Other examples for practical applications are aerial photographs overlaid by maps, X-ray vision to see inside an object or showing a movie with subtitles in two different languages.

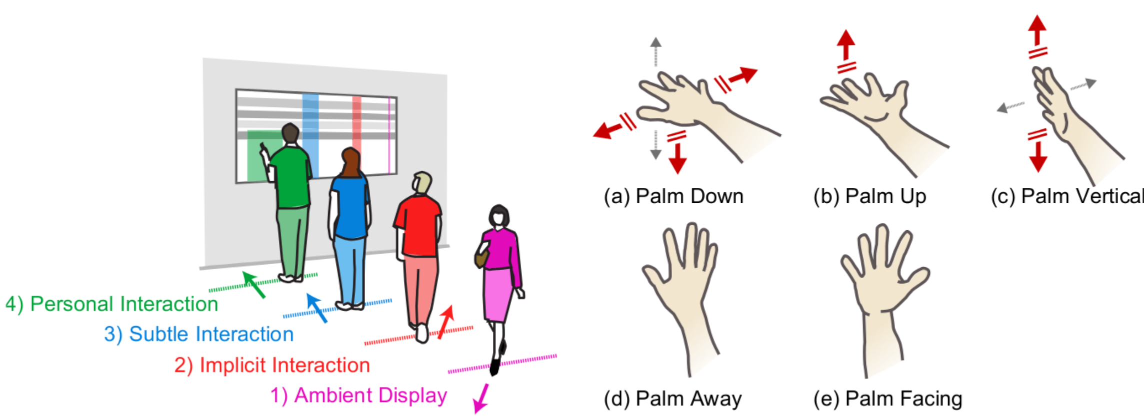
Figure 4: A satellite image augmented with a map view using the magic lens. [10]

3 Interaction Based on Direct Touch Technology and Haptics

A different approach is to detect users' movements and gestures as they are standing in front of the display. That way, the user isn't in need of auxiliary devices like mobile phones, pointing devices or any other input means. The goal is to simplify interaction, make it faster and more intuitive and at the same time save the display from wear and vandalism.

An eye gaze interaction method is described by Sibert and Jacob in [11] in which the movements of the user's eyes are recorded by a remote eye tracker and used as the basis for interaction. Moving eyes is natural, requires little conscious effort, and frees the hands for other tasks. An important side benefit is that eye position implicitly indicates the area of the user's attention. Selecting is based on dwell time, which means looking at an object for a minimum time of for example 150ms. In the year 2000, eye tracker technology was still considered immature, but today's systems or at least future systems might be decent enough to become a potential input technique standard.

Vogel and Balakrishnan developed an interactive display [1] which presents general and personal information depending on the user's distance and attention which is derived from head and body orientation. They set value on immediate usability, shared use, combining public and personal information, calm aesthetics and privacy. In figure 5(a) four different phases are defined. Phase 1 shows general information such as weather or activity levels in branch



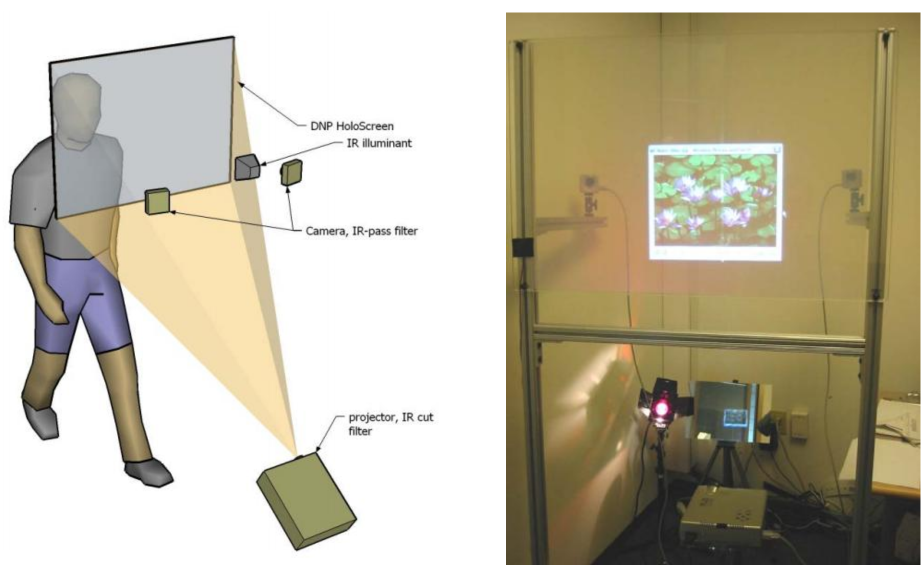
(a) Four interaction phases, facilitating transitions from implicit to explicit, public to personal, interaction.

(b) Postures and gestures used for interaction with the display.

Figure 5: [1]

offices. In the second phase the user will be notified of any urgent information and of the interactivity of the display. Phase 3 is a personalized subtle interaction phase which can be entered by standing still for a moment within a certain threshold distance. In this phase some personal information augments the display. Phase 4 is active when the user steps closer to the display. Now more personal information like private messages can be displayed safely by using small font size and exploiting natural body occlusion. Interaction is possible by using hand gestures (Figure 5(b)). In Vogel and Balakrishnan's setup a marker glove is used to track the gestures. However, in the future this tracking might be possible using only cameras and appropriate pattern recognition algorithms. If a user wants to quit interaction with the display or doesn't even want to start it, he can either leave the phase by turning around or show his palm facing the display like a stop sign.

A. Wilson is facing the issue of gestures on touch sensitive screens [12] as well as B. Tratz in his bachelor's thesis about the development of a multi-touch display [13]. Both developed displays which are sensitive to parallel touching on different parts of their surface. The system consists of a semitransparent plane, appearing to the user as the display, a projector in the back of the plane producing the optical output, an IR illuminant or respectively infrared LEDs on the sides of the screen facing into the plane and two or respectively one camera susceptible to infrared light (Figure 6(a)). A realisation of the TouchLight prototype is shown in figure 6(b). A computer obtains input by image processing and creates output data for the projector. If users touch the plane parts of the infra red light refract on their fingers, which can be detected by the camera. Using edge extraction the created image can be processed and interpreted. Thus, it is even possible to interpret different gestures, e.g. zooming in our out by applying the pinch gesture as known from the Apple iPhone, or rotating an object by pressing a finger on it and drawing a circle around it using another finger. Because of the possible distinction between touch events the screen can be used as a platform for collaborative interaction. The surface of the multi-touch screen could also be divided into different areas enabling users to interact in parallel on their own. Besides games like PONG [13], very different applications are possible. For example Wilson mentions the possibility to use the screen like a document scanner. Holding a piece of paper against the screen, would lead the camera to take a high resolution picture. The detected position, size and orientation of the page may then be used to automatically crop, straighten and reflect the high resolution scan of the document. Alternatively, the application may project an all-white graphic on the page to clearly illuminate it [12]. Another application scenario could be using the display - camera combination as an eye-to-eye video conferencing system. With most of today's conferencing systems, it is impossible to maintain direct eye contact with the person you are talking to. However this is possible with the camera positioned right behind the semi-transparent screen. Of course, this would need some changes regarding the hardware and software, like a low-reflection screen and an algorithm to subtract the image displayed by



(a) The TouchLight rear-projection setup: a semitransparent screen, two IR cameras, an IR illuminant and a projector (b) Picture of the TouchLight prototype realisation

Figure 6: The TouchLight screen - a multi-touch rear-projection display [12]

the projector in order to send a clear picture to the corresponding conference partner. The fact that a pattern recognition algorithm is able to save and identify someone's face might be two-edged. It can be very practicable if the system is used for computer supported cooperative work, maybe for a log-in authentication or automatically saving data into a personal folder. In public spaces, however, this might become an even bigger privacy issue as for example at work. Usually the camera isn't installed visible to the user of the display. Thus, one might not even notice being monitored or captured by the camera. Another application mentioned by Wilson is an augmentation of reality. Imagine one of these rear-projection displays integrated in a clothing shop window mirroring you. By tapping onto the display one could virtually try on different styles of clothes without having to change. The processing computer would just display an overlay onto your mirror image.

Similar to the technology mentioned before, *Microsoft* introduced a multi-touch device (see figure 7) in 2007, reminding of a table, which is able to recognize touch events from multiple users, connects wirelessly to devices put upon the surface and interacts with them. It allows users to manipulate digital content by the use of gesture recognition. This is also realized by a combination of hardware and software, among others consisting of infra red cameras underneath the display, wireless technology and appropriate image processing algorithms. A fascinating feature is the possibility to transfer files between devices in a very easy way. If users want to share pictures using their digital cameras or mobile phones, they only have to put those devices (which have to be capable of a wireless technology) onto the display. The *MS Surface* recognizes these devices and pairs with them. The pictures stored for example on the digital camera pop up on the display of the *MS Surface* and can be transferred to the other device e.g. the mobile phone using a dragging gesture with the user's finger.

Figure 7: Using the *Microsoft Surface* it is possible to transfer files intuitively between devices. Source: http://commons.wikimedia.org/wiki/File:Surface_table.JPG

Figure 8: A user reaches her right hand towards her right hip to access a tool. This mechanism allows for immediate eyes-free tool selection regardless of user loction in the room. [14]

Shoemaker et *al.* published 2010 their paper [14] in which a concept for interaction with very large wall displays is presented. They use coloured balls attached to the user's "key joints" (shoulders and wrists) to detect the orientation of the body via cameras. Those markers might become unnecessary in the future when computer vision progresses. Having the information about the orientation a virtual shadow of the user's body is generated and displayed on the screen. The shadow can be intuitively controlled by moving one's body. Every user carries virtual tools with him which can be accessed by pointing at the area they are stored in, such as the shoulder or hip area (Figure 8). In this way the tools are always available and can be selected without the need of a visual search or a targeting operation at the display. Private data "is stored" in the user's torso. If sharing personal contents is desired the user has to point at his stomach area to access own data. Now he can browse and select the content of interest and drag it over to the shared display or hand it over to a specific person. Another feature is the possibility to use e.g. the combination of one's arm and the other arm's hand as a 1 dimensional slider to manipulate contents. In their setup Shoemaker et *al.* used *Nintendo* Wiimote controllers as additional input devices, but they hope that other gesture recognition methods without the need of devices become practicable. They suggest an approach by Vogel and Balakrishnan in [15] similar to the hand gestures mentioned above. However, in this case passive markers are still indispensable.

In 2007 Cernekova et *al.* published a paper [16] about single camera gesture recognition. This approach is independent of markers or any other utilities. Two setups are presented. First the camera is placed 2m above the floor and

2.5m to the right of the user. In the second setup the camera is placed above the screen and focused on the user standing in front. When the user points with his fully extended arm at the screen the captured image will be analysed and key features can be extracted. In the side image key features are the top of the head, pointing finger, shoulder and the feet position whereas in the frontal image only the top of the head and the pointing hand will be interpreted as key features. The display is separated into square shaped cells. Given that a first calibration image was taken where the user points at a specific cell of a shown grid, areas that are pointed at can be inferred. Disadvantages of this approach are the low accuracy - a cell is of size 20cm x 20cm - and a determined position for the user. If those weaknesses can be overcome in the future, this approach will be an inexpensive and intuitive solution for interaction with public displays without the need of any input device.

4 Interaction Based on Combinations of Haptics and Input Devices

There are some research groups which emphasise the potential of the combination of haptics and devices. Due to the ubiquitous character of mobile phones interaction with public displays would be possible for everybody. Combining the benefits of mobile phone technology with intuitive gestures is a promising input method.

Ballagas et *al.* developed the sweep technique [6] which is based on a Bluetooth and camera equipped mobile phone. Optical flow image processing is used to interpret the change of successive images provided by the phone camera as a relative motion. This allows the mobile phone to be used like an optical mouse known from usual PC workplaces (Figure 9). While moving the phone's joystick must be pushed vertically and held to activate the movement detection. This "clutch" is advantageous for repositioning the phone - like a mouse on the desktop. The camera doesn't need to face the screen resulting in benefits like a comfortable arm position and enabling a maximum of attention towards the screen.

Figure 9: The sweep technique can be used to control a cursor on the large display like an optical mouse. [6]

In 2008 Vajk et *al.* transfered some key properties of *Nintendo*'s Wiimote to a mobile phone in order to interact with public displays [17]. The Wiimote is a controller for *Nintendo*'s game console *Wii* equipped with an accelerometer. It is able to detect both rotational and translational acceleration along three-dimensional axes. Vajk et *al.* used the inbuilt accelerometer of the mobile phone *NOKIA* 5500 to sense motion and interpret this data as input. They presented a simple game as an application example for this input possibility. People could connect via Bluetooth to the public screen's server and join the multiplayer game *TiltRacer*, a 2D game in which a car with constant speed must to be steered through a race course. Using Bluetooth, a large screen and the integrated sensor in the mobile phone the user was freed from any cables, the constraints of the limited graphic capabilities of the mobile screen and he could control his virtual car quite intuitively while standing at an arbitrary position.

5 Conclusions

In this paper a survey of interaction techniques for public displays was presented. A distinction between three general input methods was made, which are interaction based on input devices, haptics and touch technology and combinations of those methods. To conclude, the majority of currently published papers concerns interaction based on devices. Using personal devices like mobile phones is a practical approach, since nearly everybody is in possession of one and they become faster and more powerful, making them capable of complex tasks. Input devices are of special importance if the screen is out of the user's physical reach. Those techniques are often not very intuitive, though. Therefore technically unversed people will be deterred from possible interaction. Multi-touch is state of the art and still a technology of high potential. It seems to become the standard input technology for public displays in the near future. Low prices and the touch screen's intuitive input characteristic support its popularity. However touch input is inapplicable for remote interaction as well. Combinations of mobile input devices and haptics can help to overcome this problem. In the long run interaction based on haptics only might replace the mentioned input techniques, due to its very easy, intuitive and pure (no need for devices) characteristics.

References

[1] Vogel, D., Balakrishnan, R.: Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. In: Proceedings of the 17th annual ACM symposium on User interface software and technology. UIST '04, New York, NY, USA, ACM (2004) 137–146 1, 3, 5

[2] Brignull, H., Rogers, Y.: Enticing people to interact with large public displays in public spaces. Ineract 2003 (2003) 17–24 2

[3] Churchill, E., Girgensohn, A., Nelson, L., Lee, A.: Information cities: Blending digital and physical spaces for ubiquitous community participation. Communications of the ACM 2003 **47**(2) (2004) 38–44 2

[4] Pawar, U.S., Pal, J., Toyama, K.: Multiple Mice for Computers in Education in Developing Countries. In: Information and Communication Technologies and Development, 2006. ICTD '06. International Conference on. (2006) 64–71 2

[5] Izadi, S., Brignull, H., Rodden, T., Rogers, Y., Underwood, M.: Dynamo: a public interactive surface supporting the cooperative sharing and exchange of media. In: Proceedings of the 16th annual ACM symposium on User interface software and technology. UIST '03, New York, NY, USA, ACM (2003) 159–168 2

[6] Ballagas, R., Rohs, M., Sheridan, J.G.: Sweep and point and shoot: phonecam-based interactions for large public displays. In: CHI '05: CHI '05 extended abstracts on Human factors in computing systems, New York, NY, USA, ACM (2005) 1200–1203 2, 2, 9, 4

[7] Hardy, R., Rukzio, E.: Touch & interact: touch-based interaction of mobile phones with displays. In: MobileHCI '08: Proceedings of the 10th international conference on Human computer interaction with mobile devices and services, New York, NY, USA, ACM (2008) 245–254 2

[8] Cheverst, K., Dix, A., Fitton, D., Kray, C., M: Exploring mobile phone interaction with situated displays. "Pervasive Mobile Interaction Devices (PERMID 2005), workshop at Pervasive 2005" (2005) 2

[9] Boring, S., Baur, D., Butz, A., Gustafson, S., Baudisch, P.: Touch projector: mobile interaction through video. In: CHI '10: Proceedings of the 28th international conference on Human factors in computing systems, New York, NY, USA, ACM (2010) 2287–2296 3, 2

[10] Waldner, M., Kalkusch, M., Schmalstieg, D.: Optical magic lenses and polarization-based interaction techniques. In: Proc. 13th EUROGRAPHICS Symposium on Virtual Environments (EGVE'07). (2007) 61–68 4

[11] Sibert, L.E., Jacob, R.J.K.: Evaluation of eye gaze interaction. In: Proceedings of the SIGCHI conference on Human factors in computing systems. CHI '00, New York, NY, USA, ACM (2000) 281–288 3

[12] Wilson, A.D.: Touchlight: an imaging touch screen and display for gesture-based interaction. In: Proceedings of the 6th international conference on Multimodal interfaces. ICMI '04, New York, NY, USA, ACM (2004) 69–76 3, 3, 6

[13] Tratz, B.: Entwicklung eines multitouchdisplays mit schnittstelle zum qt-framework. Technical report, Technische Universität München, LDV, Munich, Germany (April 2010) 3

[14] Shoemaker, G., Tsukitani, T., Kitamura, Y., Booth, K.S.: Body-centric interaction techniques for very large wall displays. In: Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries. NordiCHI '10, New York, NY, USA, ACM (2010) 463–472 8, 3

[15] Vogel, D., Balakrishnan, R.: Distant freehand pointing and clicking on very large, high resolution displays. In: Proceedings of the 18th annual ACM symposium on User interface software and technology. UIST '05, New York, NY, USA, ACM (2005) 33–42 3

[16] Cerneková, Z., Nikolaidis, N., Pitas, N.: Single camera pointing gesture recognition using spatial features and support vector machines. 15th European Signal Processing Conference (EUSIPCO 2007) (September 2007) 3

[17] Vajk, T., Coulton, P., Bamford, W., Edwards, R.: Using a mobile phone as a "wii-like" controller for playing games on a large public display. Int. J. Comput. Games Technol. **2008** (January 2008) 4:1–4:6 4

Mobile Authentication and Security

Junjie Zhang
junjie.zhang2010@mytum.de

February 10, 2011

Abstract

Nowadays mobile security methods for protecting user data are developing quickly. Some approaches using auxiliary channels are efficient and capable in the authentication process. However, these methods are still based on separate scenario and difficult to compare and interchange. Therefore, an appropriate method is required, which can exploit any combination of authentication approaches and be applied on the current mobile devices. In this paper, I give an overview about these approaches, which are based on different scenarios and adopt various realizations. I introduce a toolkit named "OpenUAT" [1], which implements multiple auxiliary channels. In this section, three main components of OpenUAT and how these components interact among themselves are briefly explained. Additionally a summary about the advantages of this toolkit is exposited at last.

1 Introduction

Ubiquitous computing is a trend in the whole world. Nowadays people carrying mobile phones, personal digital assistants (PDAs) or tablet personal computers (Tablet PCs), can be seen almost everywhere. Those mobile devices are used frequently in business, in the educational area, or to realize variety of personal demands. Accompanied with masses of mobile applications, people, however, start to concern about another important aspect: security. The communication channel, in which the data is exchanged between devices, is wireless, i.e. in the air. In short distance, users do not know how the data transmitted in wireless environment compared with the traditional wire channel. If a eavesdropping in wire channel happens, the additional wire used by eavesdropper could be easily noted by users. Unlikely to the wire channel, Some malicious attacks in the wireless channel, e.g. man-in-the-middle attacks can be difficultly directly observed by users and severely influence the development of mobile applications. A man-in-the-middle attack can take place as long as the communication is unauthenticated. If the attack succeeds, attacker can easily pretend to be either party in the processing communication without knowing by the attacked two parties. The security of user data is thus being challenged and more secure approaches are urgently needed.

2 Overview

In order to ensure the occurent activities between mobile devices are secure, an authentication is necessary before the data exchange. So far, there are plenty of research in terms of mobile authentication and security. Some of them use the communication channel both for data communication and authentication between users; while others exploit an auxiliary (out-of-band) channel for authentication. That means except for the main communication and data-exchange channel, people can use another supplementary channel for authentication. This thought, which treats authentication and date-exchange into two individual parts, makes the independent research of different approaches realistic. The researchers aimed at various prototypes to implement process of authentication. With help of different techniques in areas of mechanics, acoustics, graphic identification and data analysis, many approaches to authenticate devices were successively invented, and some much maturer methods are being investigated. Now, I want to give you an overview of this subject by illustrating some achievements.

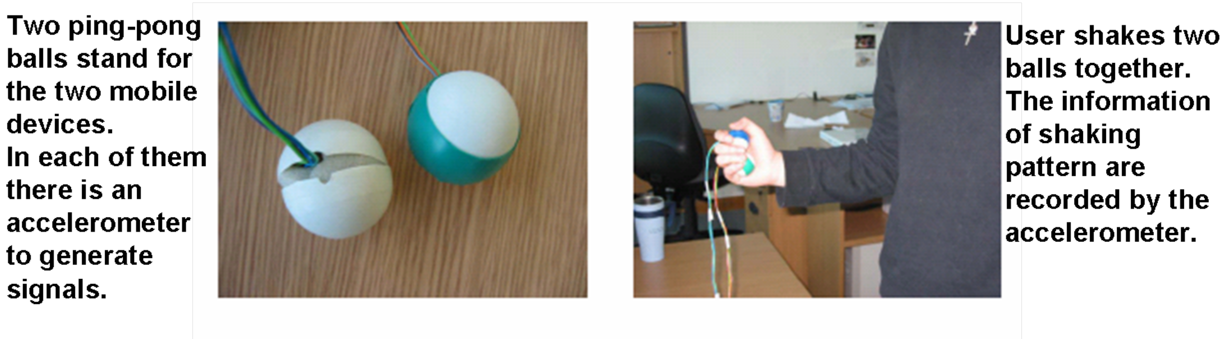


Figure 1: Devices with accelerometers and subject during data collection. Source: [2]

2.1 Motion

Here is an approach implemented by using movement pattern of devices called ShakeMe-Authentication [2], which can be realized for its lower hardware condition demand and ease of use.

As shown in Fig. 1, the two ping-pong balls stand for the two mobile devices. In each of them there is an accelerometer late on to generate signals from the movement patterns. By shaking the two "devices" together, the shaking pattern are recorded by the accelerometer. Then the data from accelerometer will be analyzed and used to create a secret key, by which the devices can establish a secure wireless channel. This approach has some advantages:

1. Shaking is very easy. People do not need to learn a complex use method.

2. The shaking pattern can be anything and the shaking time is very limited. It is basically impossible for the third party to imitate and build an interception channel, because it needs the eavesdropper to shake his device exactly in the same way as the users and he must do it in a very short time before the secure channel established.

3. When many pairs shake their mobile devices to authentication at the same time, the secret key among the pairs won't be mistaken also due to the variety of the shaking pattern, which generates different secret key.

There is another example of the gesture-based authentication method for untrusted terminals [3] (Fig. 2). When a user comes near by the wanted public terminal, the terminal detects user's mobile device. After the user affirms his device on the screen of terminal, a random shaking pattern will be chosen and displayed on the screen. Then the user performs the gesture as shown. When all the gestures are correctly executed, the terminal will grant the access right to the device and the authentication is over. This challenge-response method also leaves out the password-input process in the traditional way and successfully avoids the contact between device and terminal, which minimizes the insecurity caused by peeking and recording crime.

2.2 Visual Tag

Another way to authenticate unknown devices is by using camera-enabled devices and certain visual tag. One example of an application is based on the Ubicolor Visual Tag [4]. As shown in Fig. 3, the visual tag is colorful and divided into two parts: the upper part contains network information of the address and configuration; the lower part provides hash information. First, one of the two devices "A" uses the camera to take a picture of the visual tag shown on the screen of the other device "B". By extracting network address from the visual tag, "A" attempts to connect "B" and uses the hash information, which can be seen as an algorithm to generate a secret key. Once "B" confirms the secret key, a secret channel between two devices will be established. This approach is easy to be used as well: user just needs to take a picture, the other procedures after are automatically accomplished. The algorithm can produce abundant keys according to kinds of combination of the colorful information, which makes the key cracking by the third party impossible.

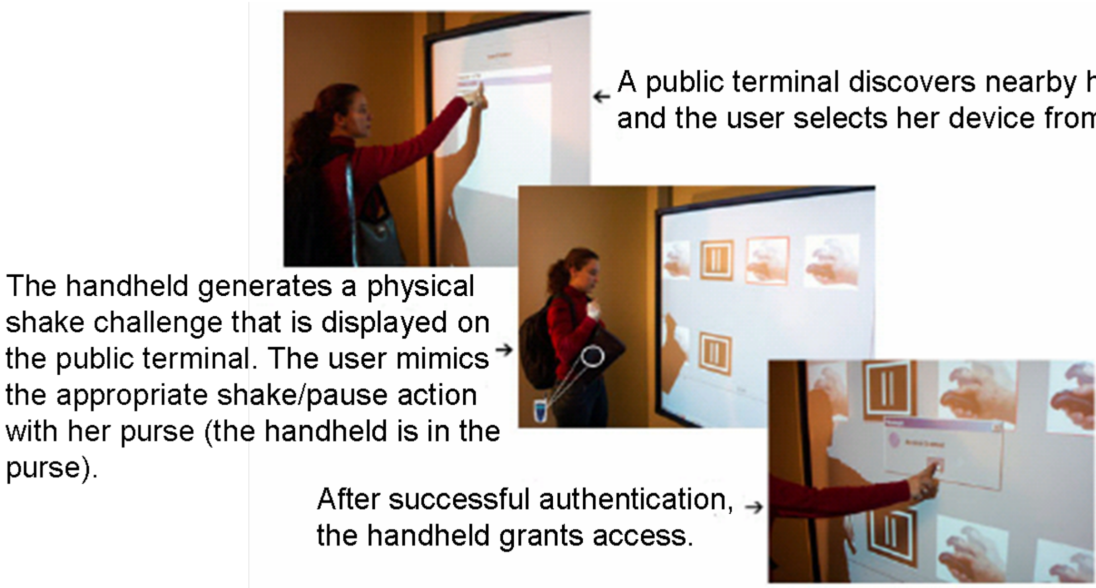


Figure 2: A user authenticating her handheld to gain access to its contents through a public terminal. Source: [3]

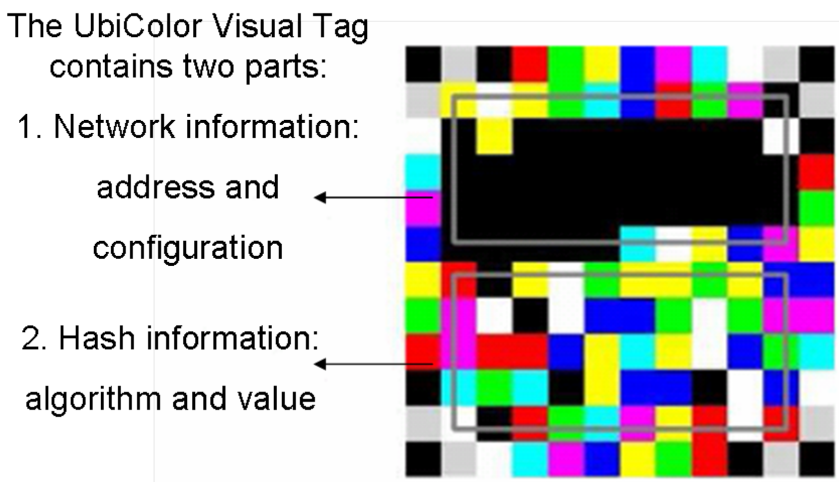


Figure 3: An Example of UbiColor Visual Tag. Source: [4]

2.3 Audio

Hearing as one of the human senses is successfully applied in the authentication area. A very simple audio authentication technique called "HAPADEP" (Human-Assisted Pure Audio Device Pairing) was presented by Claudio et *al*. [5]. This method only requests a speaker, the whole protocol thus only relies on the audio channel. The HAPADEP protocol contains two phases: the Transfer Phase and the Verification Phase (Fig. 4). First, the user starts the protocol by pressing a button on both devices simultaneously. The target device encodes the public key with the encoder and then transmits the key on the audio channel by playing the audio sequence. Because the encoded sequence has to use a fast coder in order to quickly transmit the huge information of the public key, the audio sounds unpleasant. But the whole transfer time ends quickly and the personal device then receives the audio sequence. After decoding the sequence, user obtains the public key. Now the first phase of protocol is complete (for the bilateral case, the process is similar). In next phase, each device encodes the public key with another relatively lower coding rate, and then plays the new generated audio to the user, who takes part in the authentication part and decides whether the melodies match. If they match, then the authentication qualifies, and a secret channel is established. Here, let the user be an important role in the verification phase is important. If there is man-in-the-middle attack in this case, then an unwanted audio from the attacking device would be easily discovered by user. A pleasant melody as the symbol in the end of authentication will satisfy the user as well.

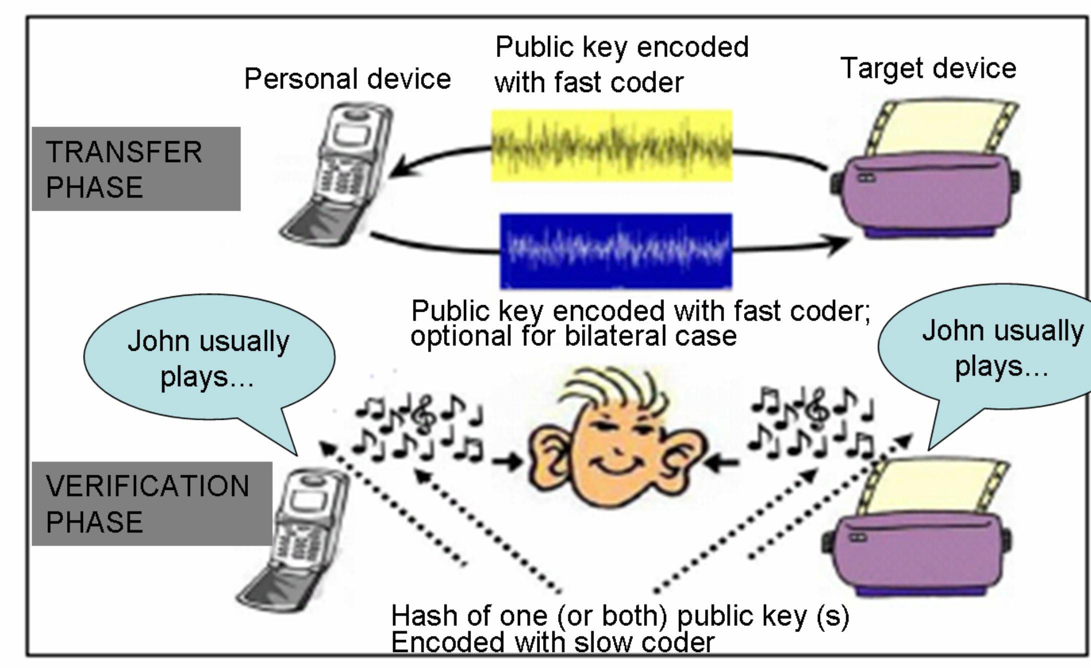


Figure 4: Implementation of Human-Assisted Pure Audio Device Pairing. Source: [5]

Here is another example of authentication application based on audio channel, however using a text-to-speech (TTS) engine to read a robust-sounding, syntactically-correct sequence, which comes from the authentication key. This whole authentication system is so called Loud and Clear (L&C) System [6]. In a simple circumstance of one-way authentication, user's device has to authenticate a target device (the bidirectional case - mutual authentication is similar because it is only an extension of unidirectional case). The target device sends his public key to the user's device over a wireless channel such as Bluetooth or infrared channel. After verifying the key on the user side, L&C generate a usually non-sensical but syntactically-correct sentence based on the string of bits, which are produced after decoding the received public key. Then the L&C system in the user device using a TTS engine broadcasts the MadLib sentence loud and clear. Now user can easily compare what he hears and what he sees displayed on the target device. If the two sentences between peer devices are the same one, authentication succeeds. However, the audio-based authentication has an obvious drawback: When the location is noisy where authentication takes place, it is difficult both in extracting useful information from received audio sequence and for user to make the decision.

2.4 Synchronized Button Presses

Different from the above approaches, this so called BEDA (Button-Enabled Device Association) [7], only requires an input button. Here, taking two cell phones as an example and adopting the simplest implementation by pressing one same number button on each cell phone (Fig. 5). The user simultaneously presses and releases the two buttons for 7 times. After each press, the user needs to wait random time to go to the next. There is a timer in each cell phone to record the elapsed time between two presses. Then each of the different elapsed time is e.g. modulo 8, that means the user obtains a 3-bits value by each button action. After 7 times user gets altogether 21 random bits value in each device, which will be then authenticated by proving them to each other in a 21-round protocol. If the authentication passes, then a secret channel will be established.

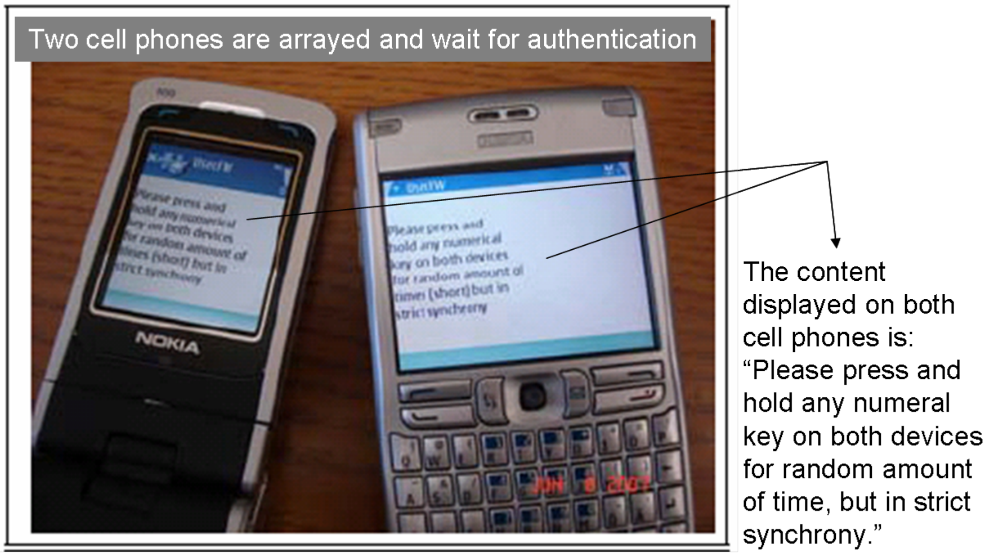


Figure 5: Simple implementation by pressing one same number button on each cell phone. Source: [7]

2.5 Ultrasound

This method makes use of spatial references between a pair of devices to create interaction and process the following authentication in spontaneous networks [8]. First step, the user's device starts to search network as well as to sense devices within its coverage. Here, the realization of sensing and spatial discovery is done with help of the Relate System [9]. The wireless relate sensors with 3 ultrasonic transducers can cover the front, left and right side of the device. They combine their own ultrasound channel with the radio frequency channel to implement this process. Then the discovered devices provide spatial parameters such as bearing and distance in order to let the host device accurately calculate their relative positions (Fig. 6). After that users take part in the authentication process. By observing the real direction, distance between the two objects, users have to choose the desired device displayed on the screen of the host device correspondingly. As following, a protocol for key exchange is implemented spontaneously. If the exchanged key proved to be reliable, a secret channel will be established between devices.

The approaches mentioned above have their own usability. However, these methods are still separate and difficult to compare and interchange. Therefore, an appropriate method is required, which can exploit any combination of authentication approaches and be applied on the current mobile devices. In the next chapter I would like to introduce a toolkit - OpenUAT [1], which can appropriately solve the problems mentioned above and provide some more functions.

3 OpenUAT

OpenUAT is a toolkit for mobile authentication. It implements multiple auxiliary channels. User can select abundant approaches and protocols, whereas developer can make full use of its feature of modular and compact to proceed on further researches. The toolkit is written in Java and can be compiled in newer java development environment (e.g. JDK versions >= 1.2.0).

3.1 Structure

From the application point of view, the components and interactions of OpenUAT is described in Fig. 7. There are three central components: `RemoteConnection`, `OOBChannel` and `HostProtocolHandler`.

- `RomoteConnection` is an interface for RF communication between devices and stochastic channels. In the case of this figure, Bluetooth, TCP sockets and some other prototypes for communication are already loaded. Developers can add the other RF channel simply by extending the abstract class `RemoteConnection`.

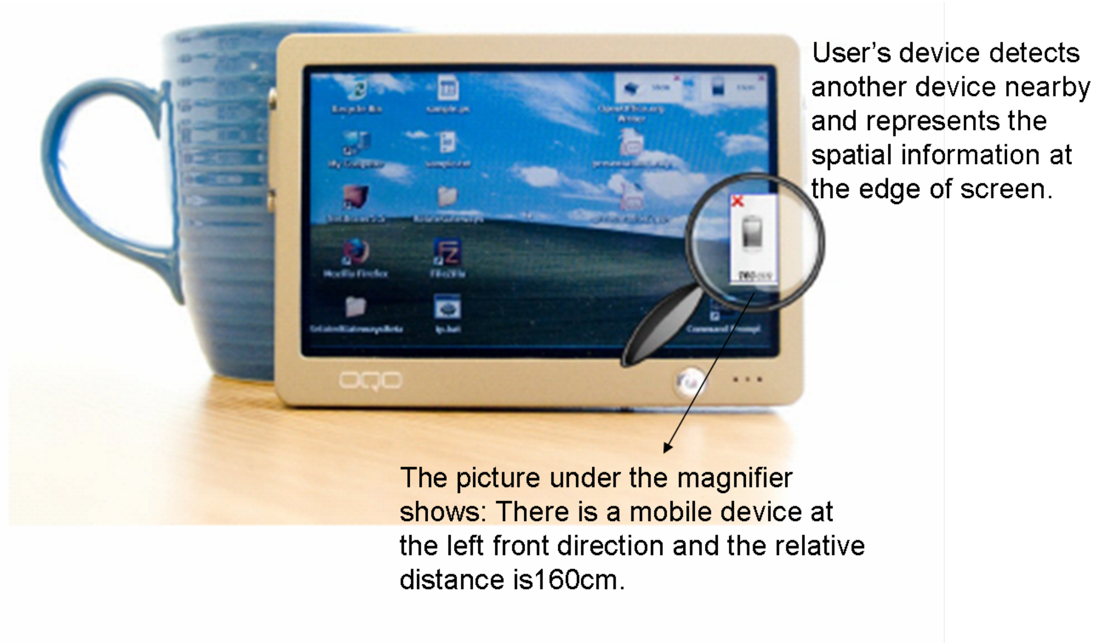


Figure 6: Near-by devices in the mobile user interface. Source: [8]

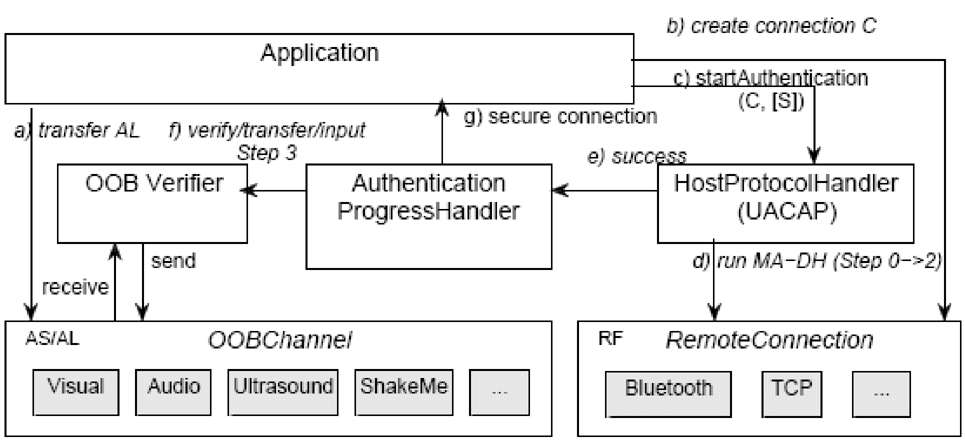


Figure 7: OpenUAT toolkit: Components and interactions. Source: [1]

- `OOBChannel` is an interface for auxiliary (out-of-band) channels, which currently includes variety of authentication prototypes based on ultrasound, motion, 2D barcodes, audio, manual string comparison, synchronized button presses and so on. For new auxiliary channel research, developers can implement the `OOBChannel` interface and its public methods `receive` and `transmit` as well as the `OOBMessageHandler` interface with its method `handleOOBMessage`. When a message is received, the `OOBChannel` notifies the `OOBMessageHandler`, which then processes the message and completes the key verification step. [1]

- The central `HostProtocolHandler` class executes steps 0 and 2 in UACAP - Unified Auxiliary Channel Authentication Protocol presented by Mayrhofer et *al.* [1]. After exchange of the DH key, it generates a homologous event with the registered `AuthenticationProgressHandler`. The different modes supported by UACAP can be pre-set before the protocol run. All other modes - the auxiliary verification using different `OOBChannel`, begin in step 3. This depends on the verification mode and the type of auxiliary channel. You can see the rough summary in Tab. 1. Sometimes the authentication process needs the user's interaction, for instance, user have to shake devises together or press a button on each device at the same time, or take part in the final decision in certain scenarios.

Channel/Mode	**Input**	**Transfer**	**Verify**
Visual		Barcode transfer	Compare sentences Manual string comparison
Audio		Audio transfer	Compare melodies Ultrasound
Motion	ShakeMe		ShakeMe
Keypad	BEDA Manual keypad entry		

Table 1: OpenUAT: Currently implemented authentication methods. Source: [1]

3.2 Advantages

Except for the functionally realization of the mobile authentication - providing approaches for creating shared secrets between two devices to prevent man-in-the-middle attacks, there are additionally advantages of OpenUAT, which are always considered since the toolkit was developed in the first place [10]:

1. Light weight: There are limited resources on mobile device compared with large-scale equipments or facilities. It is especially manifested in the hardware and applicational area, such as less memory, storage and battery-supply, slower CPU, finite communication bandwidth and I/O interface. OpenUAT is designed to be small and call resources as few as possible when authenticating. It uses static memory buffers when possible and makes the communication minimum.

2. Self-contained: OpenUAT could be compiled and ran in various devices and platforms. From this thought the toolkit contains specific libraries, which are not available in certain platforms.

3. Simple to use: From the aspect of developers, the authentication toolkit should be used easily. On the one hand, developer would abandon it in simple applications if the toolkit is hard to learn; one the other hand, faulty operation would inevitably appear when the usage is too complicated, that would then cause insecurity. Therefore, OpenUAT provides just several interfaces as a black box. Developer can easily build the secure authentication protocols without additional knowledge of the internal structure.

4. Extensible: As a toolkit, supplementary elements in the near future are unavoidable. Therefore, OpenUAT is simple to extend and will be seamless as a new whole package.

5. Vertical: There are many layers in terms of authentication context, from sensors, I/O devices, communication channel, data analyze up to user interaction. OpenUAT offers elements to realize a cross-layer action. The layer above could make use of different components in the bottom level.

6. Interoperable: The environment of ubiquitous computing is complex. Therefore, in order to complete authentication process between different devices and platforms, OpenUAT uses protocols based on primitive ASCII line or standard protocols as IETF RFCs.

Full source code is available under the terms of the GNU Lesser General Public License (LGPL) at http://www.openuat.org.

4 Conclusion

Due to the arising mobile applications, many scientists have contributed kinds of approaches to authenticate mobile devices. These approaches use various techniques based on different scenarios. They focus on diverse aspects and have their own advantages and drawbacks, thus are difficult to compare and interchange. OpenUAT as a toolkit in some way solves the problem. By providing a vast library of various approaches, it can easily be used to analyze each method or to compare the data among approaches. From the developers point of view, this toolkit is also convenient for further authentication research.

References

[1] Mayrhofer, R., Ion, I.: OpenUAT: The Open Source Ubiquitous Authentication Toolkit. month (2010) (document), 2.5, 7, 3.1, 1

[2] Mayrhofer, R., Gellersen, H.: Shake well before use: Authentication based on accelerometer data. Pervasive Computing (2007) 144–161 1, 2.1

[3] Patel, S.N., Pierce, J.S., Abowd, G.D.: A gesture-based authentication scheme for untrusted public terminals. In: UIST '04: Proceedings of the 17th annual ACM symposium on User interface software and technology, New York, NY, USA, ACM (2004) 157–160 2.1, 2

[4] Claycomb, W.R., Shin, D.: Secure real world interaction using mobile devices. In: PROC. PERMID / PERVASIVE. (2006) 2.2, 3

[5] Soriente, C., Tsudik, G., Uzun, E.: HAPADEP: human-assisted pure audio device pairing. Information Security (2008) 385–400 2.3, 4

[6] Goodrich, M., Sirivianos, M., Solis, J., Tsudik, G., Uzun, E.: Loud and clear: Human-verifiable authentication based on audio. In: Distributed Computing Systems, 2006. ICDCS 2006. 26th IEEE International Conference on, IEEE (2006) 10 2.3

[7] Soriente, C., Tsudik, G., Uzun, E.: BEDA: Button-enabled device pairing. Proc. IWSSI (2007) 443–449 2.4, 5

[8] Mayrhofer, R., Gellersen, H., Hazas, M.: Security by spatial reference: using relative positioning to authenticate devices for spontaneous interaction. In: Proceedings of the 9th international conference on Ubiquitous computing, Springer-Verlag (2007) 199–216 2.5, 6

[9] Hazas, M., Kray, C., Gellersen, H., Agbota, H., Kortuem, G., Krohn, A.: A relative positioning system for co-located mobile devices. In: Proceedings of the 3rd international conference on Mobile systems, applications, and services, ACM (2005) 177–190 2.5

[10] Mayrhofer, R.: Towards an open source toolkit for ubiquitous device authentication. (2007) 3.2