

Philosophies and technologies for ambient aware devices in wearable computing grids

Pieter Jonker^{a,*}, Stelian Persa^a, Jurjen Caarls^a, Frank de Jong^a, Inald Lagendijk^b

^a*Pattern Recognition Group, Faculty of Applied Sciences, Delft University of Technology, Delft, The Netherlands*

^b*Multimedia Research Group, Faculty of Information Technology and Systems, Delft University of Technology, Delft, The Netherlands*

Received 9 October 2002; accepted 9 October 2002

Abstract

In this paper we treat design philosophies and enabling technologies for ambient awareness within grids of future mobile computing/communication devices. We extensively describe the possible context sensors, their required accuracies, their use in mobile services—possibly leading to background interactions of user devices—as well as a draft of their integration into an ambient aware device. We elaborate on position sensing as one of the main aspects of context aware systems. We first describe a maximum accuracy setup for a mobile user that has the ability of Augmented Reality for indoor and outdoor applications. We then focus on a set-up for pose sensing of a mobile user, based on the fusion of several inertia sensors and DGPS. We describe the anchoring of the position of the user by using visual tracking, using a camera and image processing. We describe our experimental set-up with a background process that, once initiated by the DGPS system, continuously looks in the image for visual clues and—when found—tries to track them, to continuously adjust the inertial sensor system. We present some results of our combined inertia tracking and visual tracking system; we are able to track device rotation and position with an update rate of 10 ms with an accuracy for the rotation of about two degrees, whereas head position accuracy is in the order of a few cm at a visual clue distance of less than 3 m.

© 2002 Elsevier Science B.V. All rights reserved.

Keywords: Ambient aware devices; Personal digital assistants; Differential global positioning system; UMTS; Ad-hex networking

1. Introduction

The growth and penetration of sophisticated digital communication systems, infrastructures, and services, has been increasing over the last decade. Examples are: Internet, electronic mail, multimedia, pagers, Personal Digital Assistants (PDA), and mobile telephony. From marginal penetration years ago, these systems and services became a commodity in the consumer markets today. Current advances are *wireless and mobile systems* that support the communication of different media, such as data, speech, audio, video and control [4,27]. European wireless network and mobile phone services are currently centered around four available technologies: WAP, UMTS, Bluetooth, and mobile positioning systems [26].¹ Positioning systems will become an integral part of mobile phones, such that services

can be made dependent on the location of the user in the network. In future, three developments are of importance:

First, one can observe that more and more mobile phone-like devices start to include accessories such as a small keyboard, a display, and a speech interface. They are emerging as hybrids between a mobile phone and a wireless laptop personal computer or a PDA.

Secondly, we observe that computing resources are becoming ubiquitous: everywhere and available at all times. More and more consumables, durable products and services contain sensors, actuators, processing units, and (embedded) software. Integration technology makes these components smaller and more versatile.

Finally, we observe that communication and computing is becoming increasingly personal. The device is always on-line, the user is identifiable, it knows about the user's position, environment and preferences. Any service or content provider will have the opportunity to adapt existing services to the mobile terminal, and develop and provide novel end-user services. Three categories of generic

* Corresponding author.

E-mail address: pieter@ph.tn.tudelft.nl (P. Jonker).

¹ <http://www.webproforum.com/wap/>

services (middleware) for customized mobile end-user services (applications), can be distinguished:

1. *Personalization*: adapt services to the user's position and orientation (location-awareness) combined with the user's preferences, profile and current foreground and background applications (ambient awareness).
2. *QoS adaptation*: adapt services to the quality of the network (and vice versa) and the terminal of a user.
3. *Brokerage*: Broker between personal requirements and service capabilities, i.e. that negotiate and match between user profiles and service characteristics.

With higher bit rates, and, eventually, when nanocomputing becomes feasible, this integration of personal computer and personal communication device will be pushed further, eventually leading to a wearable computer system that is a valid Personal Digital Assistant, communication device, powerful computer, as well as entry point for mobile services that take the ambience of the user into account.

In this paper we will focus on technologies for ambient awareness of the future mobile computing/communication device. In Section 2 we will focus on location, context and ambient awareness of a mobile user, we will describe the possible context sensors, their required accuracies, their use in mobile services as well as a draft of their integration into an ambient aware device. In Section 3 we will focus on position sensing as one of the main aspects of context aware systems. We describe our setup for a mobile user that has the ability of Augmented Reality, which can be used indoor and outdoor in the professional field by urban planners, architects, utility (sewer) maintenance, but also by common users in guided museum tours, path finding for tourists and outdoor games. We will then focus on a set-up for pose (position and orientation) sensing of a mobile user, based on the fusion of several inertia sensors and (D)GPS. In Section 4 we focus on the anchoring of the position of the user by using visual tracking, using a camera and image processing. We describe our experimental set-up and our strategy of a background process that continuously looks in the image for known visual clues and when found tries to track them, to continuously adjust the inertial sensor system.

2. Ambient awareness in mobile communications

2.1. Ambient awareness

Ambient awareness is the process of a personal computing device in acquiring, processing and-in the background, possibly unattended by the user- acting upon application specific contextual information, taking the current user preferences and state of mind into account. Consequently, a device is context aware if it can respond to certain situations or stimuli in its environment, given

the current interests of the device and its user. One of the main topics of context awareness is location awareness; the device must know where it is. Many groups work on context awareness. Most of them implement applications like route planning, while only a few groups research the global view. We adopted the view of Ref. [19] who discussed context awareness from the following four viewpoints:

Contextual sensing. Sensing is the most basic part of context awareness. A sensing device detects various environmental states, such as position and time, and presents them to the user or to the services that want to make use of them. This sensing can be used for synthesizing context information, for instance to determine if it is dark outside.

Contextual adaptation. Using context information, services can adapt to the current situation of the user, in order to integrate more seamlessly with the user's environment. For instance, if a mobile phone enters a non-disturbance area like a theater or an area with loud noise like a disco, it can automatically disable sound signaling and only use vibration.

Contextual Resource Discovery. Using the own context of a device may be adequate for some services, but in some situations more information from the environment is needed. The device should be able to discover other resources to determine the context of other entities like persons, devices, etc. For instance, if the user wants to display a movie and his personal device cannot adequately display this, the device can look for unoccupied display screens in its neighborhood. Another example is that if the user is stuck with a problem, his device might seek a nearby adviser for his problem. In this case other context aspects are used than the current location only.

Contextual augmentation. Some services will not only adapt to the current context, but also couple digital information to the environment. Depending on the user's perspective this can be viewed as the digital data augmenting reality or reality augmenting the digital data. An example of digital data augmenting reality is tour-guiding [6] in which the device gives information about nearby attractions. An example of reality augmenting digital data is when one can view someone's current location -or other context- when one is visiting his/her homepage on the web.

In the sections below we discuss technologies that can be used for contextual sensing. We split the sensors in two parts: position sensors and others, as using position is more evident than sensors such as a heartbeat sensor.

2.2. Context sensors

Position sensors are already used in a variety of applications, also in other fields than context awareness. Many sensor types exist with different accuracy, price and range. The list below -partly taken from Refs. [14,20] and [14] is not exhaustive, but one or more technologies from this list is at hand.

Inertial sensors sense accelerations and rotations [15]. This means that they can follow changes in position. These motion, or inertial sensors are quite fast, so they can track fast motions. However, due to inaccuracies, in the inexpensive sensors, they can only track for a short period without drift, so another—usually slower—positioning system with a lower drift is needed to correct for accumulated errors. High accuracy means a high price, weight and volume. On a device for the consumer market the error will grow from 10 to sometimes 100 m within a minute.

The *(Differential) Global Positioning System* (DGPS) consists of 24 satellites around the earth. A receiver can determine its 3D position by using the information of at least four satellites. By using information from another GPS receiver with known position, the normal error of about 10 m for a commercial system, can be lowered to 1 m (DGPS). Accuracies of 1 cm can be obtained but the price for this is currently too high for the consumer market. The two disadvantages are that (D)GPS cannot be used indoors, and that it only gives the position, not the orientation, meaning that other sensors, like magnetometers, are needed to obtain orientation.

Network access points give a rough position of the device. In case of a dense network, with access points (AP) everywhere—like Bluetooth or GSM—the PDA can get a qualitative accurate position. For the current GSM network a cell ranges from 100 m to 1 km. Bluetooth cells will be around 10 m, but unlike GSM, there is no Bluetooth network available. So using a proprietary network means building and maintaining a new infrastructure.

GSM/GPRS/UMTS networks have a global network of base stations, so an inaccurate position is always available using the cell information. With multiple antennas or measuring the propagation time or signal strength [29] of the signals from different base stations one may get an accuracy of about 50 m (in small cells). The signal can also be used indoors, but sometimes this will become too weak. A strong point is that communication and localization can be done via the same network. Using the signals to get a better accuracy than using cells alone is under research, and up to now not available to the public.

Beacons, such as the GSM base stations can be used for localization because they are already there. Others, like ultrasound, radio or infrared beacons can also be used. Such a beacon grid will mostly only be used for localization and not for communication. Setting up a network of beacons could be expensive, and will probably be used only indoors, or at special locations, e.g. on bus stops. The beacon can broadcast its own position, or the device could calculate its position from one or more beacon signals, or both. Apart from the infrastructure that is needed, reflection of signals could be a problem in dense networks. This means that methods to suppress or ignore the reflections are needed, to avoid errors.

Visual markers are cheap to construct and easy to mount on walls/doors, objects, etc. These passive beacons, a.k.a. fiducials, involve a whole other technique, using cameras and

image processing. If the PDA sees a marker and recognizes it, it knows approximately its position. And in case more markers are seen at the same time or shortly after each other, the PDA could calculate an accurate position and orientation, if the exact position of the markers is known. The markers are not restricted to man-made patterns; they include pictures, doorposts, lamps or all that is already available. But detecting or recognizing natural markers is complicated, while artificial markers can be designed for accuracy and easy detection. Humans use them all the time: street names, house numbers, room numbers, etc.

Visual Models can be used to group all markers together into a model of a building, shopping center, museum, or part of a city. This model has to specify features that a camera at certain positions can see. In case of outdoor use, for example a wireframe model can be used that includes outer contours of buildings and their windows. The camera's position and orientation can be tracked by detecting features such as edges, i.e. the transitions from building to air, and matching them onto the wireframe model that was retrieved from the network. When the camera's position and orientation is determined, other objects in view may be used to keep tracking the position and orientation. The biggest challenges are building the database, selecting good properties of the features so that they can be detected in various circumstances, and coping with the various lighting conditions that are experienced both indoors and outdoors. This method of localization could be very accurate, but is very computational intensive. Calculations can be spread over the PDA and the backbone via the network link. As smart CMOS cameras will be used in similar applications in future, this system will become feasible.

Images could be used to assist the user in self-localization. If orientation is important, and the PDA only knows its position through GPS, the PDA could present key images to the user. If the user turns the PDA to the corresponding direction, the PDA could display an arrow to show the way in which to walk. This type of visual information could be used indoors and outdoors, but this requires a big database containing the images. A solution can be to generate images from a model, retrieved from a database in a Geodetic Information System.

A part of context awareness is the understanding of a bit of the world around the device. Information from the context awareness sensors can be used to automatically start certain services, like displaying a navigation map, when entering a building, or starting the electronic wallet service when near a counter. These examples can be realized by monitoring the location of the PDA, but other sensors could be useful in other services, like a light sensor to automatically turn on the backlight of the PDA in the dark.

Motion sensors are cheap sensors that only sense a change of motion. Using this information the PDA may determine what the bearer is doing: sitting, walking,

running, etc. Refs. [9,21,23] use this information to display the activity of the user.

Body sensors obtain their information from the user itself by measuring items such as heartbeat rate, skin resistance, brain activity or the position of the limbs [9]. This information can be used to determine the current activity of the user, as well as possibly parts of the user's state of mind.

PDA user activity sensors, obtain information from, e.g. monitoring a key pressed on the keyboard, or more generally monitoring the communication with the user.

A *microphone*, and pattern recognition can be used to detect many circumstances: a loud pub, a silent library, or the user talking to someone. This information could be used by a mobile phone when someone calls. Sounding an alarm is of no use if the surrounding noise is louder than the signal. Detecting other sounds may be useful, like detecting traffic, sea, or wind. These sounds may even be used to infer the location of the user [5].

The location could be used to start services like navigation or presentation of location dependent info. But from position also a social context can be derived. The PDA could make use of the fact that the PDA is in a coffee room, meeting room or other location that has a certain context. For example in the coffee room, one does not want to receive email, and in a meeting room one does not want to be disturbed at all. Another example is tour guidance, e.g. through a museum. The PDA can display information on near objects [6].

Visual sensing provides a tremendous amount of information about one's environment, and it is potentially the most powerful source of information among all contextual sensors. However, retrieving this information is very difficult, due to the fact that we cannot use the image directly but need image processing and pattern recognition techniques to extract higher-level scene information from it. Examples are detecting a person, a car, a room, or building. Useful information about the light source is easily determined by measuring the overall incoming light into a camera: In sunlight there is much red light, while in cloudy weather the light will be bluish. The PDA could use this to determine the weather conditions, or to know it is outside or inside [1,5,23]. Visual context markers (pictograms) can be used as well: coffee room, playground, toilet, museum, etc. This again presumes an infrastructure of such markers, but as these markers are simple to construct and apply -in contrast with position markers they do not need to be fixed at an exact position -they could be convenient [23].

2.3. Required accuracies for some context based services

Navigation using accurate positioning is used in navigation services. This can for example be used in a dark unknown environment, e.g. in a calamity situation. One needs a real accurate position, down to a meter for outdoor navigation and about 10° in orientation. For indoor

navigation the accuracy in position should be around 50 cm as there can be two doors within two meters.

Navigation using rough positioning can also be used in navigation services, but then in light, more known environments. This means that the device knows in what room or part of the corridor it is. The device has to give a description of the scene and where to go, such as: 'first door left', so that the user can navigate himself. The device still has to monitor the position in case the user takes a wrong turn, but the accuracy could be 10 meters. This can be viewed as an inaccurate position, e.g. about 5 meters off. However, using this definition, in case the device is near a wall separating two rooms, it cannot know in which room it is.

Super fluous Navigation using proximity sensing means that the PDA should be able to signal if one is near a person or device of interest. If the PDA really can sense a nearby device and its relative position, no absolute position is needed. However, to discriminate devices or persons quite near to each other and if the PDA cannot sense nearby devices directly, an accurate position of the PDA as well as that of the nearby device should be available to calculate the relative position. In most cases only a rough position of about a few meters is necessary.

Information presentation based on position means that information is sent only when passing by a certain location. Examples are: a shop or post-office is nearby, or person is in a room. This is possible using the rough position of the PDA. Outside the PDA has to know, e.g. a house number or the name of a street or the nearest building. When entering a building, a network change can be enough to display information about the building.

Service discovery based on position needs rough position information. Examples are Automatic Teller Machines, and printing or projection facilities within a building. If there is only a single network within a building, or worse, only one access-point, a rough position needs to be built up by the PDA and its sensors.

Communication with nearby devices is when one wants to communicate with other nearby device. Examples are payments, video/DVD rental, photo printing, showing media content on a screen, or telling your car to lock. A convenient way is to point at the device to communicate with it. Consequently, to establish the identity of the target device, there should be a very short-range link between the two devices and, moreover, a direction sensitive one. After the devices identified each other, the communication could go via normal network access. If there is no such a link to establish the identity, the position and the direction of the PDA can be used to determine which device one wants to communicate with. A simple solution is to display on the PDA all the available nearby devices and let the user make a selection.

2.4. Mapping sensors on services

For outdoor navigation, (D)GPS is a good candidate, because it has the highest accuracy. The drawback is that it

Table 1
Example of assembled context information in a PDA at time t

Actor	Location (room, address, country)	Default action	Current action	Planned action	Aware state
Me	F253 NL-2628CJ-1 (coffee room)	Working	Break	Meeting John in 5 min at F256 (meeting room)	Sitting, talking, drinking, relaxing, waiting
John	? NL-2628CJ-1	Working	?	Meeting with me in 5 min confirmed at $t-5$	?
Others	F256 NL-2628CJ-1 (meeting room)	Working	Meeting	?	Discussing
Meeting room	F256 NL-2628CJ-1	Meeting place	Meeting place	New meeting in 5 min	5 people inside, discussing

does not sense the orientation. However, this can be overcome by using a compass or the movement of the user. Another possibility is to use a camera and a model of the world. From features/markers in the image a position can be calculated and tracked, possibly in the cm range. This option will only be feasible if the calculations can be done in such a way that the device will still be wearable. For indoor navigation, GPS cannot be used. Indoors it is easier to set up a network of beacons. With ultrasonic or radio beacons good accuracies could be realized. If visual markers are used as beacons, a camera is needed. In both cases, if the position should be updated frequently, motion sensors are needed to track the PDA's position during a few seconds.

In conclusion, the position awareness has to be established in some way to detect services, devices and humans, and to make context aware sensors. To communicate with a device, the PDA has to 'discover' this device first in order to make a connection with the device. This can be done by pointing the PDA at the device.

If every device has an active or passive beacon that transmits or shows its own Globally Unique Identification (GUID), the PDA could detect this GUID. As GUID a (visually coded) IP-address or URL can be used. If the signal is weak enough and the receiver is directionally sensitive, only the GUID from the device pointed at is received. This GUID is then used to establish a connection with the device via the communication network, but of course also a dedicated infrared link could be used.

If there is an accurate positioning system available, the position and orientation of the PDA and the position of each device can be used to determine which device is pointed at. In that case the network, device and PDA should know each other's exact location.

2.5. Implementing a device with ambient awareness

Information from the context sensors and the intentions of the user are building blocks for an ambient aware device. Added to our own agenda plus the agendas of other persons, places, rooms or devices with which we interact, it could fill a distributed database, partially maintained in the device and partially at the service provider. Network services that wish to use contextual information can try to login to the database of the PDA and query this database, as far as it is allowed.

Equally so, the PDA can retrieve information via the network from service providers or other PDAs. Some information does not leave the PDA and therefore is not accessible to services not running on the PDA, while position information might be sent to the database of the network provider. This view of a contextual information database is not yet complete. Issues are: protocols for contextual information sensors, such that no specific knowledge on the sensors is needed; how and where to combine the data; how to setup the database infrastructure so that services can effectively use the databases; the protocols for querying the databases, etc.

Table 1 shows an idea of ambient awareness of a PDA that was built up from own sensor information and a person's agenda, stored in the PDA, as well as queries to other PDAs and devices over a network. The aware state is inferred from the position sensors, the other contextual sensors, as well as the agenda that describes the actions. The row Me in Table 1 is assembled within the PDA, whereas the other rows are assembled from information retrieved via the network services, via the network from other PDAs or with a direct link from other PDAs, in all cases as far as this was allowed.

3. Position sensing

3.1. The maximum set-up: augmented reality

We have chosen *visual augmented reality* as the maximum feature of potential context aware applications (Fig. 1). Augmented reality implies that the system is not only aware of the context; it also merges virtual audio-visual information with the real world. A system that anticipates the interests, intentions, and problems of the user and reacts accordingly -by merging proper virtual information with the real world -must continuously monitor the surroundings and actions of the user. As such we process information from sensors, combine it with information from other sources, and present the output in visual form to the user.

Our user carries a wearable² terminal and a lightweight see-through display (Fig. 2). In this display the user can see

² Our current breadboard version fits in a backpack.

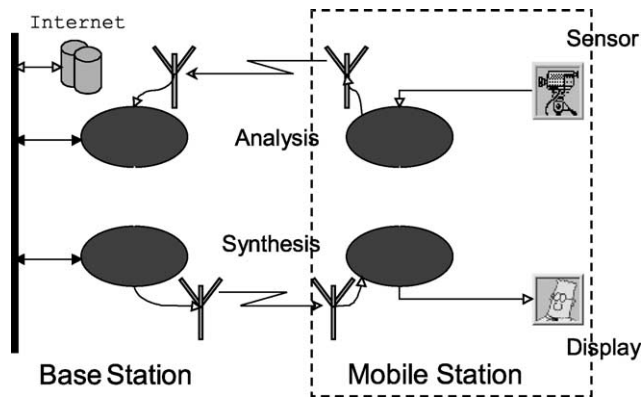


Fig. 1. A PDA using network services.

virtual information that augments reality, projected over and properly integrated with the real world. The wearable system contains a radio link that connects the user to ubiquitous computing resources and the Internet. A camera captures the user's environment, which, combined with gyroscopes, accelerometers, compass and DGPS, makes the PDA fully aware of the absolute position and orientation of the user with such accuracy that virtual objects can be projected over the user's real world, without causing motion sickness. The rendering update rate of our system is 10 ms [20]. Camera images are sent to the backbone and matched to a 3D description of the environment derived from a GIS database of our campus, to determine the user's position and to answer questions of the user and his PDA that relate to the environment.

We consider our set-up as an application-specific context aware system for future generation personal wireless communications. Although we investigate the maximum set-up, various low-cost versions can be derived, the least demanding one being a system without Augmented Reality. This lowers the requirements for the positioning accuracy and update rate drastically. In such a version the camera can be used to realize the awareness of the user's position (where is he/she?), the user's attention (what is the user looking at, pointing at, maybe thinking about?), and the user's wishes (what is the problem, what information is needed?).



Fig. 2. Augmented reality set-up.

3.2. Position sensing for context awareness

This section presents a low-cost sensor combination and data processing system for the determination of position, velocity and heading, to be used in a context aware device. We aimed at testing the feasibility of an integrated system of this type and to develop a field evaluation procedure for such a combination. Navigation on a flat and horizontal ground (e.g. humans walking around), only requires an estimation of a 2D position and a heading. However, also the height may vary slightly along the path. An inertial tracking system is only able to accurately track the orientation in 3 Degrees Of Freedom (DOF) ((ϕ, ψ, θ) = Roll, Pitch, Yaw as named in avionics or pan, tilt and heading). To make an accurate 6-DOF inertial tracking system, including positional (X, Y, Z) information, some type of range measurements to beacons or fiducial points in the environment is required. Noise, calibration errors, and the gravity field produce accumulated position and orientation drift in an inertia based system. Accelerometers and gyroscopes are very fast and accurate, but due to their drift, they have to be reset regularly, in the order of once per second. Orientation requires a single integration of rotation rate, so the rotation drift accumulates linearly with the elapsed time. Positions can be determined by using the double integration of the linear accelerations, but this makes that the accumulation of position drift grows with the square of elapsed time. Hybrid systems attempt to compensate for the shortcomings of each technology by using multiple measurements to produce robust results. Section 3.1 has described our approach. Section 3.2 presents the system overview, sensor calibration, sensor fusion and filtering. Section 3.2 presents the results and conclusions.

3.3. Position sensing approach

The inertial data are processed in a strapdown mechanization mode [7,8,24], based on the following expression for

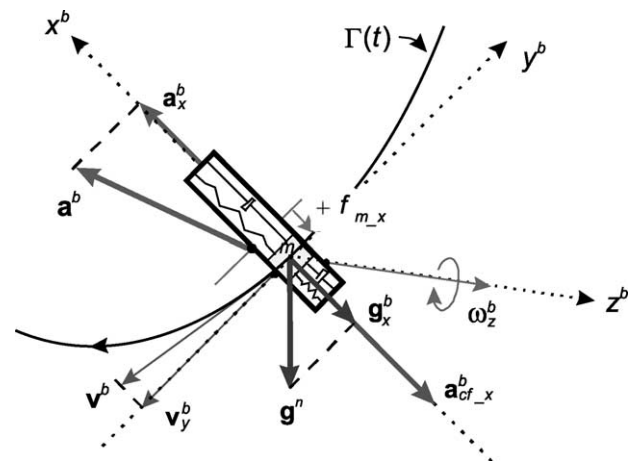


Fig. 3. Specific force as a function of accelerations along a reference system attached to a moving body (x -axis).

a one component specific force in a body reference system (see Fig. 3, that explains the forces considered, acting upon the seismic mass of the accelerometer), as a function of the linear acceleration a_x^b , the apparent centripetal acceleration a_{cf-x}^b and the corresponding axial component of the static gravitational acceleration g_x^b (the superscripts b denote the vector components in the body reference system):

$$f_{m-x} = a_x^b + a_{cf-x}^b - g_x^b. \quad (1)$$

In Fig. 3 the body is following the path $I(t)$ and turning with angular speed ω_z . An accelerometer is rigid mounted to the body on x direction, and in figure we draw all the forces that acts on accelerometer and body.

The corresponding vector form (with the specific force vector now denoted by a and the correction terms of centripetal and gravity acceleration expressed in the body coordinate system) is:

$$a^b = a - \omega \times v^b + C_n^b g^n. \quad (2)$$

with ω the angular velocity vector, v^b the velocity vector, given in the coordinate system b , and C_n^b the rotation matrix from the local coordinate system n to the body coordinate system b .

Roll-pitch-yaw angles (ϕ, ψ, θ) can be used to represent the attitude and heading of the mobile user. If the direction cosines matrix C , defining the attitude and the heading of the

the system to define the attitude and heading because they have direct physical interpretation, but this representation is not used in the differential equation. We use quaternions because they do not lead to singularities. Using quaternions, the differential equation to be solved takes the form:

$$\begin{aligned} \dot{Q} &= \frac{1}{2} Q \Omega, \text{ or } \begin{bmatrix} \dot{Q}_0 \\ \dot{Q}_1 \\ \dot{Q}_2 \\ \dot{Q}_3 \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} 0 & -p & -q & -r \\ p & 0 & r & -q \\ q & -r & 0 & p \\ r & q & -p & 0 \end{bmatrix} \begin{bmatrix} Q_0 \\ Q_1 \\ Q_2 \\ Q_3 \end{bmatrix} \end{aligned} \quad (4)$$

where $Q = Q_0 + Q_1 i + Q_2 j + Q_3 k$ the quaternion associated with the attitude of the PDA, and $\Omega = [p \ q \ r]^T$ its instantaneous angular velocity. A numerical integration method must be used to solve this equation. We use the fourth-order Runge–Kutta integration algorithm, which performs the best when compared with the rectangular or trapezoidal method. The direction cosines matrix can be expressed in terms of quaternion components by:

$$C = \begin{bmatrix} Q_0^2 + Q_1^2 - Q_2^2 - Q_3^2 & 2(Q_1 Q_2 - Q_0 Q_3) & 2(Q_1 Q_3 + Q_0 Q_2) \\ 2(Q_1 Q_2 + Q_0 Q_3) & Q_0^2 - Q_1^2 + Q_2^2 - Q_3^2 & 2(Q_2 Q_3 - Q_0 Q_1) \\ 2(Q_1 Q_3 - Q_0 Q_2) & 2(Q_0 Q_1 + Q_2 Q_3) & Q_0^2 - Q_1^2 - Q_2^2 + Q_3^2 \end{bmatrix} \quad (5)$$

user, is given, the roll-pitch-yaw angles can be extracted as follows:

$$\begin{aligned} C &= \begin{bmatrix} s_x & n_x & a_x \\ s_y & n_y & a_y \\ s_z & n_z & a_z \end{bmatrix} \\ \theta &= \arctan\left(\frac{s_y}{s_x}\right) \pm k\pi, \psi \\ &= \arctan\left(\frac{-s_z}{\cos(\theta)s_x + \sin(\theta)s_y}\right), \\ \psi &= \arctan\left(\frac{\sin(\theta)a_x - \cos(\theta)a_y}{-\sin(\theta)n_x + \cos(\theta)n_y}\right). \end{aligned} \quad (3)$$

The attitude can be determined using gyrometric measurements. This method also allows us to estimate the heading (yaw), which is not possible with the accelerometers or inclinometer. In this case, a differential equation relating the attitude and the instantaneous angular velocity has to be integrated. Roll, pitch and yaw angles are used as output of

The flow-chart of the strapdown navigation algorithm implementing Eq. (5) is presented in Fig. 4.

We neglected the g -variations and the Earth rotation rate, because of the small dimensions of the test area, of the relatively low people velocities (about 1 m/s) and of the reduced rate sensitivity of the used gyroscopes. Also we neglect the small Coriolis force acting on the moving mass as a consequence of the rotation of the inertial sensors case.

3.4. Position sensing hardware

Three sets of sensors are used: the Garmin GPS 25 LP receiver combined with an RDS OEM4000 system to form a DGPS unit, a Precision Navigation TCM2 compass and tilt sensor, and three rate gyroscopes (Murata) and three accelerometer (ADXL202) combined in one board, linked directly to a LART platform [2] developed at Delft University (Fig. 5). The LART platform contains an 8-channel fast 16-bit AD-converter to acquire synchronous data from accelerometer, gyros and in future temperature data. The latter is useful to compensate the drift due to temperature variations in the sensors. The Garmin GPS

at 16 Hz over a serial line. Therefore, when the TCM2 and the gyros are read out simultaneously, there is an unknown difference in the time of the physical events. We took the relative latency into account by attaching a time stamp to the readouts.

3.5. Sensor fusion and filtering

The goal of the sensor fusion is to estimate the angular position and rotation rate from the input of the TCM2 and the three gyroscopes. In case we need the data for Augmented Reality this position is extrapolated one frame into the future to estimate the orientation at the time the image is shown on the see-through display. At standstill, we estimate roll and pitch from inclinometer and accelerometer measurements. We use the redundant information from the accelerometer to get better precision. Roll and pitch are computed from the gravity-component in the body frame, which are directly measured by the accelerometers. The expressions of attitude angles as a function of the gravity in the body frame are:

$$\psi = -\arcsin\left(\frac{g_x}{g}\right) \quad \phi = \arcsin\left(\frac{g_y}{g^* \cos(\psi)}\right), \quad (6)$$

$$\text{or } \phi = \arccos\left(\frac{g_z}{g^* \cos(\psi)}\right).$$

To predict the orientation one frame into the future, we use a linear motion model: we add the offset implied by the estimated rotational velocity to the current orientation. This is done by converting the orientation (the first three terms of x) to quaternions and using quaternion multiplication to combine them.

For moderate head rotations (under about 100 degrees per second) the largest registration errors we observed were about 2°, with the average errors being much smaller. The biggest problem was the heading output of the compass sensor drifting with time. The output drifted by as much as 5° over a few hours, requiring occasional recalibration to keep the registration errors under control. The magnetic environment also could influence the compass error, however, for short times we can compensate this by using only the gyro readings.

4. Position anchoring by using visual tracking

The system described above, using gyros, accelerometers, compass and tilt sensor, still has a considerable drift that has to be compensated by a system that locks the device onto the real 3D world. Outside buildings, a DGPS system can be used to roughly indicate the position of the PDA. Inside buildings DGPS cannot be used. Furthermore there is a gap between the resolutions of the DGPS system and the inertia system. A system that could bridge this gap is a system that tracks beacons in the field of view of a camera.

The 3D vector of position and orientation is referred to as *pose*. Pose estimation from a camera can be used to compensate the drift of the inertia system. To recover the pose we need to have a model of the world. This model could be a description of a building in terms of the wireframes describing outer contours and contours of windows of buildings, but this could as well be a man made passive visual beacon, a.k.a. a fiducial, which is fixed onto a known position in the world. This fiducial can also specify its own pose, e.g. in a dot or bar code, knowledge that has to be retrieved from the image of the camera looking at that fiducial.

4.1. An experimental camera set-up

We are in the development of a camera system that continuously looks if it can find fiducials in the image. When it has found this fiducial, the system tracks the fiducial as long as possible. However, the system also tracks other features in the image and relates this to the tracked fiducial, such that when the fiducial is out of sight, these other features can be used to keep the tracking going in subsequent video frames. To simplify this, we investigated the matching of line pieces found in the camera image onto a wireframe model of the world. For sake of simplicity we assume a camera mounted on a human's head, looking slightly downward, and we have chosen for a self-localization error of 10 cm and 5°. We tested our system on a soccer field with green grass and white lines. Currently we are in the phase of augmenting this 4-DOF ($(X, Y, \psi, \theta) = (X, Y, \text{heading}, \text{tilt})$) set-up to a full 6-DOF system for a hand-held PDA with camera, as well as tests in more realistic scenes.

A pose must now be found that yields the best match of the 2D image from the camera with the 3D model of the world. Two approaches are possible: matching in image-space and matching in world-space; [18] surveys both methods. If the camera pose and internal parameters are known, features from one space can be projected into the other space. When matching in world-space [3,12,17] the movement of the world-projected image features is directly determined by the camera pose, but the error in the projected image features is dependent on the distance to the camera. Furthermore, the image features have to be found first, which is usually costly. When matching in image-space, the problem is that the pose itself is difficult to describe using the movement or position of the image features. However, the verification of a predicted pose could be fast, because we only need to find the features shift from an expected position in the image to get an error measure, and not the features themselves.

In our set-up, we take the best of both approaches: we find approximate poses by matching in world-space, and increase accuracy by verifying in image space. Of course, this means that we have to deal with the problems for matching in world

space. To determine the distance from the features to the camera, the perspective transformation causes the range measurements to be very sensitive to camera tilt and pixel noise, especially in features that lie close to the camera's horizon. Although the camera's tilt relative to the human user might be fixed and calibrated, the camera's tilt relative to the world will be influenced by bumps and vibrations (e.g. when a person starts walking), leading to bad matching results. To attack this problem, we use two techniques. First, we attach an inclinometer to the camera, so that the camera's instantaneous tilt can be determined whenever an image is grabbed. Second, we measure the features with subpixel accuracy.

Speed is of some importance. It is therefore natural to adopt a two-tiered approach, where a local search in image space tracks the human's pose at real-time speeds, and a slow global search in world space runs as a background process, verifying the local search's results, and re-initializing the local search when it fails. Refer to Fig. 6 for an illustration of the description in the Sections 4.2 and 4.3.

4.2. Global search

The global search transforms measured lines in the image to coordinates relative to the camera, and, using a model, generates a number of candidate poses, which are verified and graded in both world and image space.

The first task of global search routine is to detect lines in the image. A first choice for this would be to apply edge detection on the input image followed by a global Hough transform [13] on the edge image. However, as we are dealing with (a) a relatively high radial lens distortion, (b) the presence of the curved lines in the field of view, and (c) a required subpixel accuracy, we divided the line segment search into two steps: (a) finding line segments, and (b) accurately determining the line segments' positions. To find line segments, we divided the image into 8×6 subimages of 40×40 pixels—small enough to assume that all lines are mostly straight. Since the first step needs to give a coarse result only, we use the Hough transform (HT) of the edge

image E, while taking only a small number of bins for the orientation and distance (θ, ρ). Using the best two peaks in the HT of each subimage, we run sub-pixel edge detectors along lines perpendicular to the lines found with the HT, and thus find a number of edge positions with subpixel accuracy. We then convert the edge positions to the calibrated camera frame, by feeding the edge positions in the lens distortion calibration formula (removing the lens radial distortion and skew of the image plane axes), and then fit lines through points from one parent line, with least squares and leave-out.

Knowing the field lines in the calibrated image (Fig. 8, left), we then convert the lines to human relative coordinates using the camera's known pose relative to the human head (Fig. 8, right). Since we are dealing with a structured environment, we determine the main orientation ϑ of the line segments, and proceed by matching the projection on the main orientations of the field with the projection of our measured lines on their own main orientation. The 2D matching problem is thus reduced to two 1D matchings, reducing the complexity by one order. For the main orientation ϑ and the orientation perpendicular to it, ψ , we find a number of candidate positions, which follow from minima of a matching cost function (Fig. 9). We combine candidate positions on the two axes in an exhaustive way, and calculate the match between the observed line pattern and the projection of the translated and rotated field model (given the candidate position and main orientation).

The (image) line matching function calculates the perpendicular distance of the center point of each measured line to all model lines whose orientations are close to the measured line's orientation, and take the smallest distance as the distance. A penalty score is then generated from the distance by feeding it into a sigmoid function, using the line segment's estimated distance from the camera as normalizing constant. The candidate camera pose's score is the sum of all the penalty scores. A very forgiving threshold is set (linear to the number of points), to remove only the worst pose candidates, and the remaining candidates are used as input candidates for the local search step.

4.3. Local search

The local search takes an estimate of the human's pose, calculates the corresponding field line pattern using a model, and matches that with the measured lines. If not only the estimate of the human's pose is fed into the local search, but also the estimated parameters plus or minus a small increment, we get (for three parameters, x, y, ϕ) $3 \times 3 \times 3 = 27$ possible line patterns. The correct parameters are simply those that yield the line pattern that matches best with the measured image. Alternatively, we can determine the lines offset from the expected position as a function of small pose changes (Image Jacobian [16]), and solve the optimal pose change in least squares sense. Generating the expected line

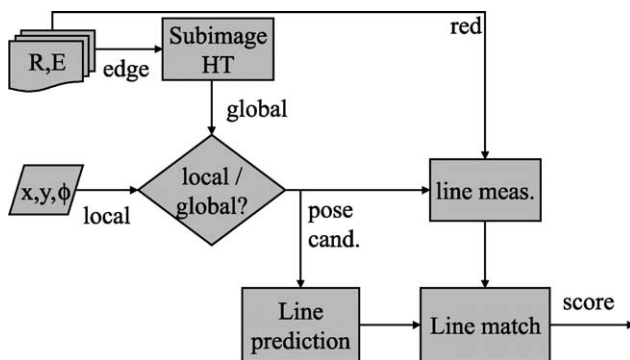


Fig. 6. Data flow of the global search. R and E are the red component and edge-label images.

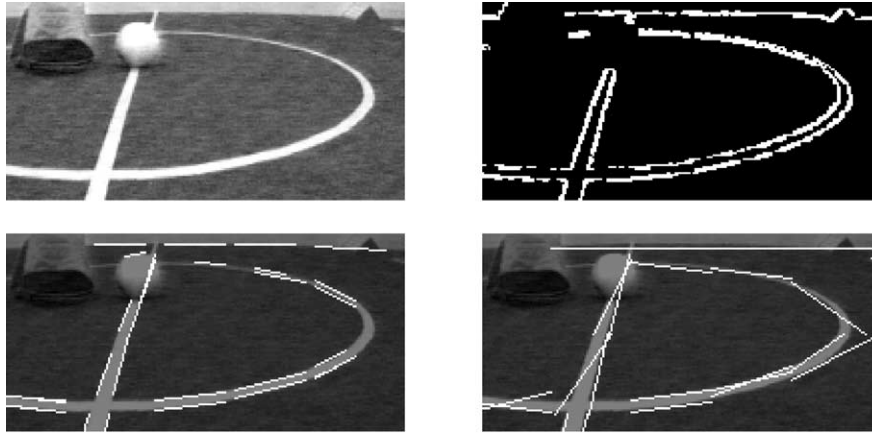


Fig. 7. Clockwise, starting left/top: Original image (red component), the label image indicating field-line edges, the coarse Hough's lines, the accurate measurement.

pattern is implemented using standard image formation methods [10,11], using the camera tilt measure given by the inclinometer, and off-line calibration information about the camera internal parameters [25,28].

The size of the increment is determined by the expected error of the inertia sensing system. For each expected line pattern, we look for line segments in a local neighborhood in the image, using the image formation (local search) output in the exact same way as the output of the coarse Hough transform (global search). In some locations, the lines will not be found (due to occlusion or other errors), so we will find a subset of the expected line segments, displaced slightly due to the error in our prediction. We then generate, for each pose candidate a matching score, using the same line-matching method as described above but with equal weights for all measured lines. The candidates with the best score are used to reset the inertia sensing system. If none of the candidates generates a good score, the global search is activated again.

4.4. Calibration and experiments

To be able to generate an accurate measure of what the camera would see, given a particular pose, we need to calibrate the camera's internal parameters well. We use the Zhang [28] algorithm, which requires three non-coplanar views of a planar calibration grid (a checkerboard pattern in our application) to estimate principal point, image axis skew, pixel width and height, and lens radial distortion. We also calibrate the color space with an interactive program that (for now) allows the user to adapt manually to the current lighting conditions.

We tested the line detector for the global search on real images (see Fig. 7). The line detector has the statistics shown in Table 2 for synthetically generated images. The subpixel edge positions are found by calculating the second derivative in a 5-pixel neighborhood and finding the zero-crossing.

We also ran the line segment finding algorithm on test images, Figs. 7–10 show the subsequent steps taken in the algorithm. Note the improvement between Fig. 2, right/bottom and left/bottom; the accurate line detector also clips off the line where it no longer satisfies the linearity constraint. The global search usually gave about 20 candidates, which was reduced to about half by forgivingly matching in world coordinates.

The matching cost—see Section 4.2—on Fig. 9 left has local minima where the lines in the back of the image match with goal line, goal area line, top of center circle, center lines, etc. The matching cost on the right shows minima for the lines in the left/top of the image matching with the left of the field, the left line of the goal area, and the left side of the center circle.

Both global and local self-localization methods show good results on test images as well as in a real-time environment. The algorithm used to determine a matching cost deserves some further discussion. Since the scene is dynamic, many parts of the field will not be visible due to occlusion or image noise. Therefore, in matching the expected and measured line set, we take the lines measured, and try to find support in the expected image, instead of the other way around. Although lines may erroneously be detected as field lines (and therefore introduce errors when matched with the field model), we assume that it is more

Table 2

Bias and standard deviation for the line estimator for a horizontal edge, and two slanted edges with increasing additive Gaussian noise (line segments up to 32 pixels long)

	Horizontal, $\sigma = 3.8$	Slanted, $\sigma = 3.8$	Slanted, $\sigma = 6$
Bias	0.02	0.02	0.02
Stdev	0.04	0.04	0.06

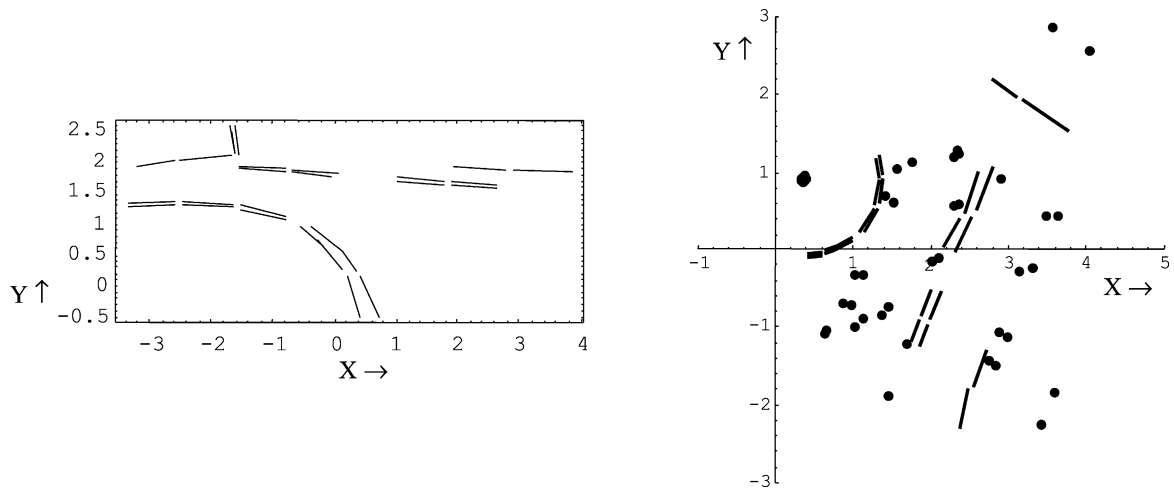


Fig. 8. Left, the measured lines after lens correction (calibrated camera coordinates). Right, the lines in robot coordinates (m), and their votes for a center circle.

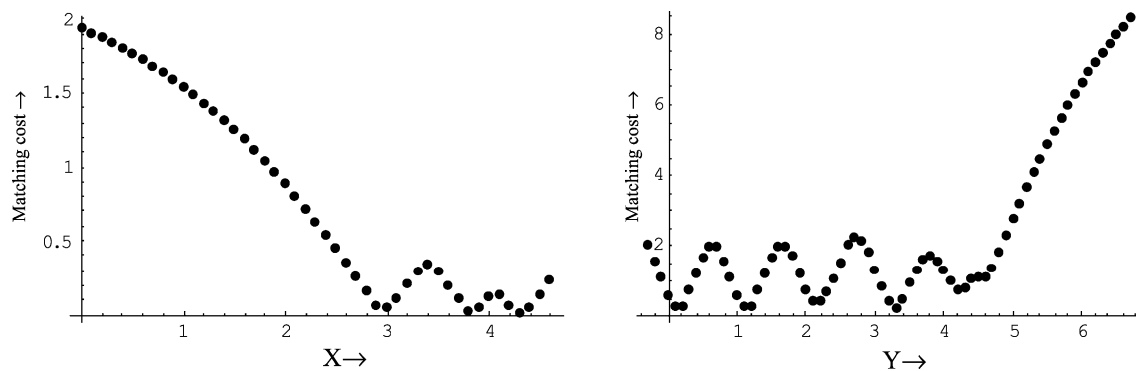


Fig. 9. Matching cost (arbitrary units) along length of field (left) and along width of field (right) in m.

often the case that field lines will not be detected due to occlusion.

The global search routine assumes zero knowledge about the human's position, and can therefore work only when 'enough' lines are visible. In future, knowledge can be inserted derived from the DGPS system or from cell

information from the communication network. The definition of enough depends on the line orientation in world coordinates and on the amount of noise, but generally speaking, we need at least one line segment in two perpendicular directions. When these lines are not available, the global search will yield too many candidates. In these cases, the human must make movements to increase the chance of observing these lines.

5. Conclusions

We described technologies for ubiquitous computing and communication. We described ambient awareness as the acquiring, processing and acting upon application specific contextual information, taking the current user preferences and state of mind into account. We described that a device is context aware if it can respond to certain situations or stimuli in its environment, given the current interests of the device. A main topic of context awareness is location awareness; the device must know where it is. We have focused on technologies for ambient awareness of a future mobile computing/communication

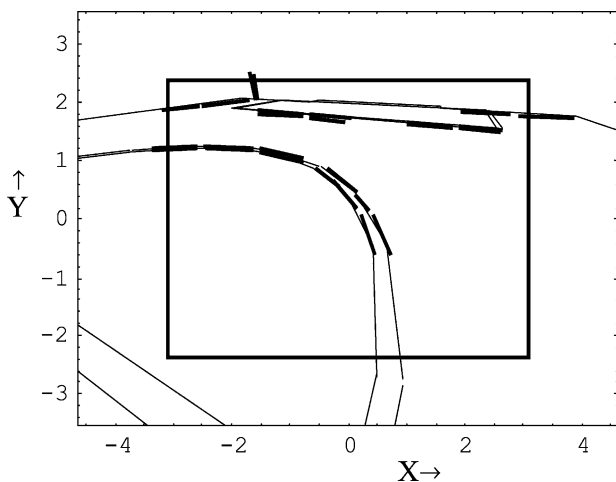


Fig. 10. Best match found. Measured lines overlaid on the expected scene (calibrated image coordinates).

device; on location, context and ambient awareness of a mobile user, describing the possible context sensors, their required accuracies, their use in mobile services as well as a draft of their integration into an ambient aware device. We then focused on position sensing as one of the main aspects of context aware systems. We described our setup for a mobile user that has the ability of Augmented Reality, which can be used indoor and outdoor by the professional as well as the common users. We then focused on a set-up for pose sensing of a mobile user, based on the fusion of several inertia sensors and (D)GPS. We described the anchoring of the position of the user by using visual tracking, using a camera and image processing. We described our experimental set-up with a background process that continuously looks in the image for visual clues and -when found- tries to track them, to continuously adjust the inertial sensor system. Within the text we described some results of our inertia tracking system as well as our visual tracking system. The inertia tracking system is able to achieve tracking of head rotations with an update rate of 10 msec with an accuracy of about 2°. The position update rate is guaranteed by the inertia system and hence also 10 msec, however, its accuracy depends on the accuracy of the visual tracking system and this was found to lay in the order of a few cm at a visual clue distance of less than 3 meter.

Acknowledgements

This work has been funded by the Delft Interfaculty Research Center initiative [DIOC] and the Telematica Research Institute.

References

- [1] H. Aoki, B. Schiele, A. Pentland, Realtime personal positioning system for a wearable computer, The Third International Symposium on Wearable Computers, Digest of Papers (1999) 37–43.
- [2] J.D. Bakker, E. Mouw, M. Joosen, J. Pouwelse, The LART Pages, Delft University of Technology, Faculty of Information Technology and Systems Available Internet: <http://www.lart.tudelft.nl>, 2000.
- [3] T. Bandlow, M. Klupsch, R. Hanek, T. Schmitt, Fast Image Segmentation, Object Recognition and Localization in a RoboCup Scenario, In: M. Veloso, E. Pagello, H. Kitano (Eds), Robot Soccer Worldcup III, LNCS Vol. 1856, 174–185.
- [4] D. Bull, N. Canagarajah, A. Nix, Insights into Mobile Multimedia Communications, Academic Press, New York, 1999.
- [5] B. Clarkson, A. Pentland, Unsupervised clustering of ambulatory audio and video, Proceedings, IEEE International Conference on Acoustics, Speech, and Signal Processing 6 (1999) 3037–3040.
- [6] N. Davies, K. Cheverst, K. Mitchell, A. Efrat, Using and determining location in a context-sensitive tour guide, Computer 34 (8) (2001) 35–41.
- [7] R. Dorobantu, Field Evaluation of a Low-Cost Strapdown IMU by means GPS, Ortung und Navigation, 1/1999, DGON, Bonn.
- [8] J.A. Farrell, M. Barth, The Global Positioning System and Inertial Navigation, McGraw-Hill, New York, 1999.
- [9] P.D. Biemond, J. Church, J. Farrington, A.J. Moore, N. Tilbury, Wearable sensor badge and sensor jacket for context awareness, Proceedings of the Third International Symposium on wearable Computers, San Francisco, (1999) 107–113.
- [10] O. Faugeras, Three-Dimensional Computer Vision, MIT Press, Cambridge, 1996.
- [11] J.D. Foley, A. van Dam, S.K. Feiner, J.F. Hughes, Computer Graphics, Principles and Practice, second edition in C, Addison-Wesley, London, 1996.
- [12] J.-S. Gutmann, T. Weigel, B. Nebel, Fast, accurate and robust self-localization in the RoboCup environment, Proceedings of the Third International Workshop on RoboCup (1999) 109.
- [13] R.M. Haralick, L.G. Shapiro, Computer and Robot Vision, Addison-Wesley, Reading, MA (ISBN 0-201-10877-1), 1 (1992–93) 578–588.
- [14] J. Hightower, Location systems for ubiquitous computing, Computer 34 (8) (2001) 57–66.
- [15] J.R. Huddle, Trends in inertial systems technology for high accuracy AUV navigation, Proceedings of the 1998 Workshop on Autonomous Underwater Vehicles, AUV'98 (1998) 63–73.
- [16] S. Hutchinson, G. Hager, P. Corke, A tutorial on visual servoing control, IEEE Transactions on Robotics and Automation 12 (5) (1996) 651–670.
- [17] L. Iocchi, D. Nardi, Self-localization in the RoboCup environment, Proceedings of the Third International Workshop on RoboCup (1999) 115.
- [18] F.M. Carlos, U.L. Pedro, A localization method for a soccer robot using a vision-based omni-directional sensor, Proceedings of the Fourth International Workshop on RoboCup (2000) 159.
- [19] J. Pascoe, Adding generic contextual capabilities to wearable computers, Proceedings of the Second International Symposium on Wearable Computers October (1998) 92–99.
- [20] S. Persa, Tracking Technology, Sensors and Methods for Mobile Users (GigaMobile/D3.1.2), December 2000.
- [21] C. Randell, H. Muller, Context awareness by analysing accelerometer data, The Fourth International Symposium on Wearable Computers (2000) 175–176.
- [22] A. Ronald, B. Hoff, H. Neely III, R. Sarfaty, A motion-stabilized outdoor augmented reality system, Proceedings of IEEE VR'99, Houston, TX March (1999) 252–259.
- [23] A. Pentland, B. Schiele, T. Starner, Visual contextual awareness in wearable computing, Proceedings of the Second International Symposium on Wearable Computers (1998) 50–57.
- [24] D.H. Titterton, J.L. Weston, Strapdown inertial navigation technology, IEE Books, Peter Peregrinus Ltd, UK, 1997.
- [25] R.Y. Tsai, A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses, IEEE Journal of Robotics and Automation RA-3 (4) (1987).
- [26] The Future Mobile Market, UMTS Forum (<http://www.ums-forum.org>), March 1999.
- [27] E.K. Wezel, Wireless Multimedia Communications: Networking Video, Voice, and Data, Addison-Wesley, New York, 1998.
- [28] Z. Zhang, A Flexible New Technique for Camera Calibration, <http://www.research.microsoft.com/~zhang/calib/>, 2000.
- [29] L. Zhu, J. Zhu, Signal-strength-based cellular location using dynamic window-width and double-averaging algorithm, 52nd Vehicular Technology Conference, IEEE VTS Fall VTC 2000 6 (2000) 2992–2997.



Pieter Jonker received the B.Sc. and M.Sc. degrees in Electrical Engineering from the Twente University of Technology in 1977 and 1979, and a Ph.D. in Applied Physics at the Delft University of Technology in 1992. From 1980 he worked at the TNO laboratory of Applied Physics. In 1985 he became assistant professor and in 1992 associate professor at the Pattern Recognition Group of the Department of Applied Physics. He was visiting scientist and lecturer at the ITB Bandung Indonesia in 1991. He was coordinator of several large multidisciplinary projects - including EU projects- in the field of Robotics and Computer Architecture. He was chairman of the IAPR TC on special architectures for Machine Vision and he is a fellow of the IAPR since 1994. His research area is soft- and hardware architectures for embedded systems that include machine vision. His current focus is on grids of wearable, ambient aware devices for communication and computing, and autonomously soccer playing robots. He is the coach of the Dutch team - "Clockwork Orange" - that consists of research groups and robots from the Delft University of Technology, the University of Amsterdam and Utrecht University.



Stelian Persa was born in Cluj-Napoca, Romania on December 11, 1970. He received the B.Sc. and M.Sc. degrees in electronic engineering and telecommunications from the Technical University of Cluj-Napoca, Romania, in 1995 and 1996. From 1996 to 1998 he worked as Assistant Professor at Technical University of Cluj-Napoca, teaching Image Processing and Television. In that period he was also a local coordinator in two international projects. From 1998 he is

Ph.D. student at the Pattern Recognition Group, Faculty of Applied Sciences of the Delft University of Technology, The Netherlands on the subject of Ubiquitous Communication. His current research interest includes real time image processing, robot vision, 3D vision, and multi-sensor fusion. stelian@ph.tn.tudelft.nl



Jurjen Caarls was born in Leiden, the Netherlands on July 29, 1976. He received the M.Sc. degree in Applied Physics from the Faculty of Applied Sciences of the Delft University of Technology, the Netherlands. His M.Sc. thesis on "Fast and Accurate Robot Vision" for the RoboCup robots of the Dutch Soccer Robot team "Clockwork Orange" won the award of the best M.Sc. thesis of the Applied Sciences faculty in the year 2001. He is currently a

Ph.D. student at the Pattern Recognition Group on the GigaMobile project. His current research interest includes: 3D vision, robot vision, Robot Soccer, real time image processing, and sensor fusion.



Frank de Jong was born in Gorinchem, the Netherlands on February 15, 1973. He received the M.Sc. degree in Applied Physics from the Faculty of Applied Sciences of the Delft University of Technology, the Netherlands. He is currently working part-time with the start-up company In3D in the area of range imaging applications, while finishing his Ph.D. thesis on high-speed visual servoing, done at the Pattern Recognition Group and the Philips Centre for Manufacturing Technology. His current research interests include range imaging, real-time image processing and visual servoing.



Inald Lagendijk received the M.Sc. and Ph.D. degrees in Electrical Engineering from the Delft University of Technology in 1985 and 1990, respectively. He became Assistant Professor and Associate Professor at Delft University of Technology in 1987 and 1993, respectively. He was a Visiting Scientist in the Electronic Image Processing Laboratories of Eastman Kodak Research in Rochester, New York in 1991, and a visiting researcher at Microsoft Research

Beijing, China, in 2000. Since 1999 he has been Full Professor in the Information and Communication Theory Group of Technical University of Delft. Prof. Lagendijk is author of the book Iterative Identification and Restoration of Images (Kluwer, 1991), and co-author of the books Motion Analysis and Image Sequence Processing (Kluwer, 1993), and Image and Video Databases: Restoration, Watermarking, and Retrieval (Elsevier, 2000). He has served as associate editor of the IEEE Transactions on Image Processing, and he is currently area editor of Eurasp's journal Signal Processing: Image Communication. Prof. Lagendijk was a member of the IEEE SP society's Technical Committee on Image and Multidimensional Signal Processing. At present his research interests include signal processing and communication theory, with emphasis on visual communications, compression, analysis, searching, and watermarking of image sequences. Prof. Lagendijk has been involved in the European Research projects DART, SMASH, STORit, DISTIMA, and CERTIMARK. He is currently leading several projects in the field of wireless communications, among which the interdisciplinary research program Ubiquitous Communications at TU-Delft.