

Title		1 (83)
D0.00-v01	Confidentiality Level: CO (Confidential)	2009-04-15



Smart Objects for Intelligent Applications

Description and Assessment of Interaction Tool Set

Legal disclaimer

The material contained herein is Confidential Information which may only be used in accordance with the terms of the SOFIA Project Consortium Agreement (PCA).

Access Rights needed for the execution of the SOFIA project are granted unless excluded in Annex IV of the PCA. No other licenses to any related IPR are implied.

The material is protected by copyright laws, and may not be reproduced, distributed or otherwise exploited in any manner without the prior permission of the rights holders.

The contributors are not liable for the use of this material except as stated in the SOFIA PCA.



Title		2 (83)
D0.00-v01	Confidentiality Level: CO (Confidential)	2009-04-15



Title		3 (83)
D0.00-v01	Confidentiality Level: CO (Confidential)	2009-04-15

Change History

Version	Date	Description	Affected Sections

List of Contributors

Participating Entity	Contributing Individuals
Philips Research	Boris de Ruyter
Conante	Stefan Rap
Philips Research	Richard van de Sluis
Philips Research	Steffen Pauws
University of Bologna	Alessandra Costanzo
University of Bologna	Piero Zappi
University of Bologna	Luigi Distefano
Eindhoven University of Technology	Jun Hu



Title		4 (83)
D0.00-v01	Confidentiality Level: CO (Confidential)	2009-04-15

Table of Contents

1	Introduction.....	7
2	Interaction Paradigms	8
2.1	Hypermedia.....	8
2.2	AnthropomorphicUI.....	8
2.3	Dialogue-based interaction	10
2.4	Pen-based Interaction	10
2.5	EnactiveUI.....	11
2.6	HapticUI.....	13
2.7	Virtual Reality	15
2.8	Augmented Reality	17
2.9	Wearable Computing	18
2.10	Brain Computer Interface.....	19
2.11	WIMP: Windows, Icons, Menus, Pointer	20
2.12	One-handed Interaction.....	22
2.13	Form-based Interaction.....	23
2.14	Direct Function Control Mapping.....	24
2.15	Ambient Interaction.....	25
2.16	Tangible User Interface.....	27
2.17	Zooming user interface.....	30
2.18	Cultural Computing.....	32
2.19	Distinction of explicit vs. implicit interaction.....	33
3	Interaction Technologies.....	34
3.1	Introduction.....	34
3.2	Modality perspective.....	34
3.3	Visual.....	35
3.3.1	Input.....	35
3.3.1.1	Video scene analysis.....	35
3.3.1.2	Visual object categorization.....	37
3.3.1.3	Facial expression recognition.....	41
3.3.1.4	VideoBasedGestureRecognition.....	43
3.3.2	Output.....	43
3.3.2.1	TextDisplay.....	43
3.3.2.2	Menus.....	43
3.3.2.3	PieMenus.....	44
3.3.2.4	2D and 3D Graphics.....	45
3.3.2.5	Augmentation.....	45
3.3.2.6	SituatedDisplays.....	46
3.3.3	Input/Output.....	47
3.3.3.1	Direct Manipulation.....	47
3.3.3.2	PickAndDrop.....	49
3.3.3.3	CrossingUI.....	50
3.3.3.4	Camera Pose Estimation for Augmented Reality.....	50
3.4	Auditory.....	52
3.4.1	Input.....	52

Title		5 (83)
D0.00-v01	Confidentiality Level: CO (Confidential)	2009-04-15

3.4.1.1	AutomaticSpeechRecognition.....	52
3.4.1.2	Computational auditory scene analysis.....	53
3.4.2	Output.....	55
3.4.2.1	SpeechSynthesis.....	55
3.4.2.2	Nonspeech Audio.....	55
3.5	Haptic (touch and proprioception).....	56
3.5.1	Input.....	56
3.5.1.1	Typing.....	56
3.5.1.2	Alternative Text Input.....	56
3.5.1.2.1	Methods utilizing numeric keypad.....	56
3.5.1.2.2	Variants of qwerty.....	57
3.5.1.2.3	Chorded text entry.....	57
3.5.1.2.4	Penbased.....	57
3.5.1.2.5	Technologies available in the consortium.....	58
3.5.1.3	Pointing by Mouse or Pen or Finger.....	59
3.5.1.4	Writing, Handwriting recognition, gesture recognition.....	59
3.5.1.5	SwitchesDialsLevers.....	59
3.5.1.6	Directional input / Joysticks, Cursor keys, Jogdials.....	59
3.5.1.7	Gloves.....	60
3.5.2	Output.....	62
3.5.2.1	Tactile feedback	62
3.5.3	INPUT/OUTPUT.....	65
3.5.3.1	Smart Objects.....	65
3.5.3.2	Force feedback.....	69
3.6	Taste.....	70
3.7	Smell.....	70
3.7.1	Output.....	70
3.7.1.1	Fragrance generation.....	70
3.8	Thermoception - feeling temperature.....	71
3.9	Nociception - feeling pain.....	71
3.10	Equilibrioception -- feeling balance.....	71
3.10.1	Input.....	71
3.10.1.1	Tilting.....	71
3.10.1.2	Gesture recognition.....	72
4	Sensors and enabling technologies	73
4.1	Ultrasound.....	73
4.2	Passive IR.....	73
4.3	Active IR and laser scanning.....	77
4.4	GPS.....	77
4.5	Head movement trackers.....	77
4.6	Accelerometers.....	77
4.7	Pressure sensors, load cells.....	79
4.8	Environmental sensors.....	80
4.9	Contact sensors.....	80
4.10	Device usage logging.....	80
4.11	Physiological measurements.....	80
4.12	RF-ID	80
4.13	Blind object identification	82



Title		6 (83)
D0.00-v01	Confidentiality Level: CO (Confidential)	2009-04-15

1 Introduction

This document is the first deliverable of the Work Package 4 on User Interaction within the SOFIA project. This work package develops interaction concepts as embedded devices that support user interaction techniques and user interaction metaphors for both a variety of different smart environment situations and a variety of different users regarding individual preferences and characteristics.

Interaction in the context of smart environments can be either active (explicit) interaction or passive (implicit) interaction by means of ambient sensors and contextual data tracking. This work package will additionally consider information presentation to the user regarding her/his current situation and individual preferences. Such feedback loops will increase information awareness and hence also enhance end user acceptance and control of the smart environment. One major application focus of this work package will be on system maintenance and end user configuration, where the user should be able to configure and adjust system behaviour by means of intuitively usable interfaces. This kind of interfaces will abstract technology in a semantic fashion. Thus handling of information and technology will be easy and feasible for a large range of different people. This work package seeks to develop methods and paradigms implemented in smart embedded interaction devices that will dramatically change the interaction of users with technology.

Additional interaction paradigms will be implemented that move today's device / function oriented device interaction to a more user goal and result oriented interaction paradigm. The implicit goal of research of this work package will be the definition of the multimodal interaction- architecture (MMI architecture) that unifies the interaction devices, the multimodal components for fission and fusion of atomic interaction events and the transformers.

This report provides an overview of available interaction technologies and tools. It starts with a chapter on the basic user *interaction paradigms* which describe the archetype or conceptual model describing the principles according to which the interaction between user and system takes place. Each interaction paradigm is illustrated by a concrete example. Chapter 3 forms the core of the document and lists all the *interaction technologies* structured by modality. In Chapter 4, an overview is given of sensors and enabling technologies. These are the technologies which are not developed in the first place to enable user interaction but can be considered to be a key enabling technology for one or more interaction technologies.

This report will be used as starting point for the analysis of SOFIA's application scenarios as developed in the other Work Packages and should be instrumental in selecting the most suitable interaction techniques and technologies for realization and implementation in a later stage of the SOFIA project.

2 Interaction Paradigms

2.1 Hypermedia

The historical roots of hypertext, hypermedia and hyperlinks are generally attributed to Vannevar Bush's seminal article 'As we may think' from 1945 (<http://www.ps.uni-sb.de/~duchier/pub/vbush/vbush-all.shtml>) in which he envisions a system called 'memex' that holds a vast number of documents and photographs (storage would be realized by means of the then popular chemical microphotography), and that allows easy access to related documents by projection and an annotation mechanism and browsing history by means of so-called trails. In a way Bush has foreseen many aspects of what we know now as the World Wide Web.

More recently and on practical computing systems, the first widespread hypertext system is considered the HyperCard system delivered with Apple's Macintosh computers in the late 1980ies. Here, users can organize their knowledge in cards that have links to further cards, thus creating a mesh of information chunks, connected by the links. It is this system that is considered as inspiration for Tim Berners-Lee when he set up his vision of a globally connected version, the 'world wide web' back in 1989.

From a user perspective, access to information is done by following links in the text (in the case of Hypermedia also pictures and the like) to related or more elaborated information. Typically a link can be discerned from unlinked text by typographic means. As the paradigm is so widespread through the success of the World Wide Web, aspects of it have been integrated in other paradigms as well, such as the WIMP paradigm (e.g. links, browsing history in Microsoft Explorer).

2.2 AnthropomorphicUI

The key point of an anthropomorphic user interface is the fact that the system is to some extent rendered like a human being (greek anthropos meaning human and morphe meaning form or shape).

It is considered desirable as the computer system can be interacted with more naturally, just as if the user would be interacting with a human being. For a user interface to be considered anthropomorphic, it doesn't matter if the rendering is also visual or not, for example a speech dialogue system that you communicate with through a telephone line would still be considered anthropomorphic.

The promise of a more natural interaction is however still often hindered by the deficiencies of today's user input analysis and system synthesis capabilities, (such as speech recognition or natural language interpretation errors, prosodically monotonous or unrealistic speech synthesis, unrealistic posture or movements of the visually rendered anthropomorphic user interface agent). The human-like appearance can give the wrong impression to the user that the system is capable of understanding or acting to the same degree as a human being which is often (or still) technically not possible.

Although fully anthropomorphic user interfaces are still primarily in the domain of research prototypes, also commercially available devices or services already deploy to some degree the paradigm, as is exemplified by the voice guidance of car navigation systems (the voice is often even given a person's name by the marketing department) or by interactive voice response systems such

as for automatic train schedule information or seat reservation over telephone lines through computerized voice dialogue.

A few years ago Philips Research created the concept of an on-screen animated character in the shape of a dog, which has been developed to facilitate voice control in a home entertainment system. The idea is that such an on-screen character can help to make the interaction between user and system more social. It can prevent people from feeling uncomfortable while talking to a system, and it can act as a clear focus point on the screen. Furthermore, the shape of a dog was chosen because the capabilities of a dog are rather comparable to those of voice control systems.



Figure: Some screenshots of “Bello”, an animated character facilitating voice control

Further examples for anthropomorphic user interfaces have been developed in the course of the Projects EMBASSI and SmartKom (1999-2003). Both projects allowed also for multimodal interaction in where the users can combine spoken commands like "record that" with pointing gestures that identifies the objects for recording.



References

- Diederiks, Van de Sluis (2001), “Bello”, An Animated Character Facilitating Voice Control, INTERACT '01, Tokyo
- Christian Elting, Stefan Rapp, Gregor Möhler, Michael Strube (2003) “Architecture and Implementation of Multimodal Plug and Play” In: Proceedings of the 5th International Conference on Multimodal Interfaces (ICMI-PUI '03), November 5–7, pp. 93–100, Vancouver, B.C., Canada.
- Wahlster, W. 2006 Smartkom: Foundations of Multimodal Dialogue Systems (Cognitive Technologies). Springer-Verlag New York, Inc.

2.3 Dialogue-based interaction

Dialogue-based interaction can be considered as a special form of an AnthropomorphicUI where the interaction is primarily by means of voice input and output.

We typically can identify several system components in dialogue based systems:

an automatic speech recognition (ASR) component that analyses the users voice signal to a sequence of words or a word hypothesis graph

- [NaturalLanguageUnderstanding](#) component that analyses the sequence of words or word hypothesis graph from the ASR and produces a syntactical and semantic interpretation thereof
- [DialogueManager](#) that maps semantic interpretation of user input to system actions and maps system events and dialoge turns, clarification questions and the like to semantic structures for output
- [NaturalLanguageGeneration](#) component that maps the semantic structures from the dialogue manager to surface representation (i.e. a sequence of words, possibly augmented by syntactic structure)
- [SpeechSynthesis](#) component that produces an intelligible audio signal from a word sequence (and possibly also from syntactic structure) of the NaturalLanguageGeneration

Depending on the breadth of the modeled domain, it is sometimes possible to omit some of these components. For example, if implementing a tightly restricted and carefully crafted command and control language for a limited domain, mapping from speech recognition to semantics or even system actions can sometimes be a trivial one-to-one mapping, or a dialogue manager might directly concatenate speech waveform snippets (thus omitting NaturalLanguageGeneration and SpeechSynthesis). However, in doing so this makes it often difficult or impossible to extend the system towards a multimodal system where information given verbally is combined with pointing gestures to objects or presented graphically. Thus it is beneficiary to make even trivial mappings explicit in order to ease integration of multimodality at a later stage.

2.4 Pen-based Interaction

The pen based interaction paradigm assumes that users are doing most of their interactions by using a stylus.

The paradigm has become widespread with personal digital assistants (PDA) and later with tablet PC, although the usage of the technology is known already for a long time. A related technique, the lightpen, utilized to interact on cathode ray tubes, has been described back in the early 1960ies (Sutherland63). Also for large displays such as interactive whiteboards, the paradigm is widely used.

A central aspect of pen-based interaction is that all activities are done by stylus, because switching between stylus and another interaction device, a keyboard, say, is inconvenient (picking up/stowing away the stylus). In consequence, the user must be able to also enter text. This is generally done by handwriting recognition in the pen-based interaction paradigm.

While for some time, pen-based interaction could be considered more or less as WIMP-style interaction plus the handwriting recognition, recently crossing-based interfaces (AccotZhai2002,

ApitzGuimbretiere2004) have become more important in research, showing that pen-based interaction must be seen independent of the WIMP paradigm. Part of this is because there are subtle differences between Mouse and Stylus. For example, while basically all touchscreen technologies can detect a hitting or stroke of the pen that is taken as a mouse click or mouse drag, many can not detect where the pen is, when it does not quite hit the display (also called hovering).



Figure: A handheld personal digital assistant employing the pen-based paradigm

References

I. Sutherland (1963) A Man-Machine Graphical Communication System, PhD thesis, Massachusetts Institute of Technology, January 1963 PDF

Accot, J. and Zhai, S. (2002) More than dotting the i's --- foundations for crossing-based interfaces. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Changing Our World, Changing Ourselves (Minneapolis, Minnesota, USA, April 20 - 25, 2002). CHI '02. ACM, New York, NY, 73-80. DOI= <http://doi.acm.org/10.1145/503376.503390>

Apitz, G., Guimbretiere, F. (2004) CrossY: A crossing based drawing application. UIST 2004, 3-12.

2.5 EnactiveUI

Enactive User Interfaces (EI) is a relatively young field of research and describes user interfaces that heavily rely on the senso-motoric skills of a user rather than symbolic or iconic knowledge. The promise of enactive interfaces is that of a utilization of movement patterns and motion dexterity learned before in the real world - in order to make operating computerized or other systems more efficient (or easy or powerful).

As psychologist Bruner described in 1968, there are three possible types of knowledge used when interacting with the world: symbolic, iconic and enactive. While most interaction paradigms are stressing the first two, enactive interfaces are stressing the third. Enactive UI are often related to tangible UI, but there are tangible UI that are not enactive (i.e. those stressing symbolic or iconic knowledge,).

One implied benefit is that once you have learnt the enactive knowledge -- stored in your muscle memory, so to say -- you ought to reproduce it very easily also after years, and also without too much (symbolic, iconic) mental effort. Examples for that are walking or riding a bike or keeping a car on the road -- once you have learned it, you will always be able to recall it. Although there is undoubtedly mental effort involved, users can in general very easily utilize their symbolic and iconic processing capacity while in such an enactive interface operation. Also note that in traditional interface technologies, quite some portion of enactive knowledge is used already, such as in ten-finger-typing, handwriting recognition and gesture recognition.

So, while the promise of EI is more intuitive and possibly more engaging interaction, it has to be shown if EI are general enough to be suitable for a wide range of tasks or whether they will be confined to be suitable for rather specific settings only.



Figure: Illuminating Clay is an example of an enactive interface. While users model a clay landscape on the table (e.g. to try out different slopes for a road), the changing geometry is captured in real time by a laser scanner. The captured data is fed into an analysis, the output of which is projected back on the clay (e.g. to highlight areas with increased landslide potential). Users can directly utilize their motor skills rather than specifying things symbolically on a computer screen.



Figure: The Segway PT is an electric energy powered vehicle that is operated by slanting towards the intended direction. leaning forward results in an acceleration, leaning backward decelerates/halts. Leaning to the left results in a left turn etc.

References

Bennett P., O'Modhrain, S. (2007) Towards Tangible Enactive-Interfaces, ENACTIVE'07, Grenoble, November, 2007

Piper, B., Ratti, C., and Ishii, H. (2002) Illuminating clay: a tangible interface for landscape analysis, CHI, pp. 355-362, NY, 2002

2.6 HapticUI

Haptic user interfaces can have a broader and a more specific meaning.

In a broader meaning, all user interactions comprising the haptic modality can be considered a haptic interface, that would, for example also include typing on a keyboard, or positioning a pointing device. Almost all input technologies in human computer interaction involve the haptic modality.

In a more specific meaning, HapticUI relates to interface devices that can exert forces on the user, and specifically those that exert forces in a computer-controlled fashion. Examples for such devices

are force-feedback joysticks, vibrotactile mice, the Phantom haptic device (and others), or the vibration alarm of a mobile phone. While pressing a key on a keyboard is definitely accompanied by a force feedback by the spring loaded mechanism, we consider typing on a keyboard not a HapticUI in the specific sense.

When talking of HapticUI as an interaction paradigm, we assume the more specific meaning.

Regarding haptic user interface controls, we can differentiate between devices that mainly operate on the 'sense of touch' (signaled from nerves in the outer skin) and those operating on the 'kinesthetic sense' or 'sense of proprioception' (signaled from nerves in the muscles and joints).

Regarding the sense of touch, most of the devices rely on vibrotactile stimuli, that is from a spinning motor with an excentric mass vibrations are coupled into the shelf of the device that can be felt (examples are vibrotactile alarm in mobile phones or the iFeel haptic mouse from Logitech/Immersion). Also gamecontrollers with haptic feedback fall in general into this category.

Regarding the sense of proprioception, in general haptic devices exert the force by means of electric motors coupled to the control. Thus for instance a dial can have a feedback force directing against the movement or along the movement of the user's hand. This is accomplished by applying a current according to the intended force that the user should feel. By reapplying the right forces for each position and motion speed, different sensations can be aroused for the user -- feeling a single notch as with a stereo amplifier's balance control, feeling a fixed number of notches and hard stops at the end as in input source selectors, a dry friction, a viscous friction etc. An example of such a haptic dial is the 'iDrive' included in some high end BMW cars.

There are also generic haptic devices used in research that work on the sense of proprioception mainly as well, for example the SensAble Phantom. They work like 'inverted' robot arms, that is a user is moving around the robot's endpoint (using a thimble or the tip of a pen held by a gimbal) and the robot simulates forces according to a virtual simulated area. These kind of devices are mainly used for medical training or general 3D simulations that also simulates collisions with or deformations of objects, but also for efficiently creating virtual models for computer aided manufacturing, e.g. for the jewelery industry.

Finally, HapticUI has a role also in teleoperation, where it is used to signal force feedback to remote operators.



Image: The image shows a haptic interface, in this case consisting of three Phantom haptic devices. They are used to simulate another haptic device, a shape-changeable dial (Michelitsch&al2004). Users can feel haptic properties of the simulated shape-changeable dial through the thimbles. The robot arms exert forces onto the thimbles in accordance with the modelling. In that way, it is possible do design controls in virtual and make experiments to find the right force strength etc. before actually making the first prototype.

References

Georg Michelitsch, Martin Osen, Jason Williams, Beatriz Jimenez, Stefan Rapp (2004) "Haptic Chameleon", Proceedings EuroHaptics 2004, June 5–7, Munich.

2.7 Virtual Reality

"Virtual reality (VR) is a technology which allows a user to interact with a computer-simulated environment, whether that environment is a simulation of the real world or an imaginary world.

Most current virtual reality environments are primarily visual experiences, displayed either on a computer screen or through special or stereoscopic displays, but some simulations include additional sensory information, such as sound through speakers or headphones. Some advanced, haptic systems now include tactile information, generally known as force feedback, in medical and gaming applications. Users can interact with a virtual environment or a virtual artifact (VA) either through the use of standard input devices such as a keyboard and mouse, or through multimodal devices such as a wired glove, the Polhemus boom arm, and omnidirectional treadmill. The simulated environment can be similar to the real world, for example, simulations for pilot or combat training, or it can differ significantly from reality, as in VR games. In practice, it is currently very difficult to create a high-fidelity virtual reality experience, due largely to technical limitations on processing power, image resolution and communication bandwidth. However, those limitations are expected to eventually be overcome as processor, imaging and data communication technologies become more powerful and cost-effective over time. Virtual Reality is often used to describe a wide variety of applications, commonly associated with its immersive, highly visual, 3D environments. The development of CAD software, graphics hardware acceleration, head mounted displays, database gloves and miniaturization have helped popularize the notion. In the book *The Metaphysics of Virtual Reality*, Michael Heim identifies seven different concepts of Virtual Reality: simulation, interaction, artificiality, immersion, telepresence, full-body immersion, and network communication. The definition still has a certain futuristic romanticism attached. People often identify VR with Head Mounted Displays and Data Suits." (wikipedia)

Example: Eat me and Drink me

At Eindhoven University of Technology, an interactive installation named ALICE is created to encourage people in Western culture to reflect on themselves, based on the narrative of 'Alice's Adventures in Wonderland' which address issues such as logic, rationality, and self. The 3rd stage of this installation is called "Eat me and Drink me", implemented using a CAVE (Bartneck and Hu et al, 2008).

Alice enters a dark corridor with many doors, which are all locked. She approaches a glass table on which a small golden key lays. She uses the key to open a tiny door that leads to a garden, but Alice is too tall to enter. She approaches the table again, and this time she notices a bottle labeled "Drink Me" and later a little cake labeled "Eat me". By drinking from the bottle she shrinks and by eating the cake she grows. Eventually she manages to have the appropriate size to enter through the tiny door.

The installation tries to manipulate the visitor's relative size in comparison to the environment by using a 5 side CAVE. The visitor entered the CAVE and had the impression to stand in a virtual room. A cookies box labeled "Eat Me" and a bottle labeled "Drink Me" are placed on top of a small table. When the visitor drinks from the bottle, the virtual room enlarges, giving the impression that the visitor is shrinking. When eating the cookie, the virtual room shrinks, giving the visitor the impression that he/she is growing.

The floor of the cave is equipped with pressure sensors that allow us to determine the visitor's position in the CAVE. Depending on his/her location, the perspective of the projection is adjusted to give a true 3D impression of the virtual room. The bottle features touch and tilt sensors to detect drinking. The cookie box is equipped with a microphone that allows us to detect the visitor's chewing sounds when eating the cookie.



Figure: 'Eat and drink me' installation

References

Bartneck, C., Hu, J., Salem, B.I., Cristescu, R., Rauterberg, G.W.M. (2008). Applying virtual and augmented reality in cultural computing. *International Journal of Virtual Reality*, 7(2), 11-18.

2.8 Augmented Reality

"Augmented reality (AR) is a field of computer research which deals with the combination of real-world and computer-generated data (virtual reality), where computer graphics objects are blended into real footage in real time. At present, most AR research is concerned with the use of live video imagery which is digitally processed and augmented by the addition of computer-generated graphics. Advanced research includes the use of motion-tracking data, fiducial markers recognition using machine vision, and the construction of controlled environments containing any number of sensors and actuators." (wikipedia)

Augmented reality (AR) is an interaction paradigm providing the user with self-explanatory information that is spatially coherent with the observed scene. This is achieved by augmenting in real-time a video stream captured by a camera with virtual graphical objects that are properly aligned with the world 3D structure as well as contextually close to user needs.

The following pictures refers to a sample AR application that has been recently developed by the University of Bologna and the Italian Aerospace Research Centre (www.cira.it) within a joint research project dealing with aeronautical servicing called ARIS (Augmented Reality to Increase Safety). As illustrated by the pictures, the developed context-aware AR system acts as a virtual assistant providing the user with real-time self-explanatory graphical information on the pre-flight maintenance procedure concerning several parts of a Cessna (<http://www.cessna.com/>) airplane (i.e. the cockpit and the engine oil tank).



Figure: Context Aware Augmented Reality application

2.9 Wearable Computing

"Wearable computers are computers that are worn on the body. They have been applied to areas such as behavioral modeling, health monitoring systems, information technologies and media development. Wearable computers are especially useful for applications that require computational support while the user's hands, voice, eyes or attention are actively engaged with the physical environment. Wearable computing is an active topic of research, with areas of study including user interface design, augmented reality, pattern recognition, use of wearables for specific applications or disabilities, electronic textiles and fashion design. Many issues are common to the wearables, mobile computing, Pervasive computing, Ambient intelligence and ubiquitous computing research communities, including power management and heat dissipation, software architectures, wireless and personal area networks" (wikipedia).

Example: Thad Starner's eyeglasses

Starner's research group, the Contextual Computing Group, focuses on projects to develop applications and interfaces for the computer to be aware of what the user is doing and to assist the user as appropriate. Several current projects at the research stage are envisioned to work together to assist a user in routine tasks such as automatically scheduling an appointment, re-directing an urgent phone call appropriately based on the user's schedule and current activity, and recognizing that the user is engaged in conversation and would prefer to take the phone call later.

"The display in your eyeglasses might also integrate a camera so the computer can see as you see."



References

<http://www.innovations.gatech.edu/wearable/>

<http://www.cc.gatech.edu/~thad/>

2.10 Brain Computer Interface

A brain computer interface (BCI) is a direct communication pathway connecting the human brain to a computer. BCI research follows three major goals (Dornhege, 2007):

1. To provide a new communication channel for patients with severe neuromuscular disabilities bypassing the normal output pathways, towards neuroprosthetics applications that aim at restoring damaged hearing, sight and movement.
2. To provide a powerful working tool in computational neuroscience to contribute to a better understanding of the human brain.
3. To provide a generic novel independent communication channel for human-computer interaction.

At Eindhoven University of Technology, van Aart et al designed an EEG headset for neurofeedback therapy. to achieve enjoyable neurofeedback therapy in the home environment (van Aart, Klaver, Bartneck, Feijs, & Peters, 2008).



Figure: EEG headset for neurofeedback therapy.

References

Dornhege, G. (2007). *Toward brain-computer interfacing*: MIT Press.

van Aart, J., Klaver, E., Bartneck, C., Feijs, L., & Peters, P. (2008). EEG HEADSET FOR NEUROFEEDBACK THERAPY. Paper presented at the Biosignals - International Conference on Bio-inspired Signals and Systems.

2.11 WIMP: Windows, Icons, Menus, Pointer

WIMP was first demonstrated in 1962 (Engelbart, Friedewald, Institute, & Alliance, 1962). Tracing back to as early as around 1970, the WIMP paradigm is still the dominant user interaction paradigm for computers and especially for PCs and workstations. The name originates from the utilized entities: Windows, Icons, Menus, Pointer.

Operation is done mainly by utilizing a general pointing device, typically a mouse, to access and manipulate information on the screen by pointing, clicking and dragging. The scarce screen real estate is managed by overlapping windows, minimizing content to icons, and by accessing a multitude of functions through (typically hierarchically nested) menus.

As virtually all PCs sold today, irrespective of an underlying Windows, Apple or Linux operating system, utilize the WIMP paradigm, and through its omnipresence, WIMP can be considered as a known interaction paradigm by most if not all users nowadays.

There are usage scenarios and situations where choosing a WIMP approach is highly questionable,

for example for operating an entertainment system in a car, or operating a mobile handset etc. It can be stated that WIMP requires a sufficiently focused (rather than intermittent) work style as is typical for operating a PC, with a high degree of visual attention to what is happening on screen -- this is not the case e.g. for driving a car, where operating the entertainment system is a secondary task, and the visual sense of the driver is already almost completely occupied by the primary task. Maneuvering the pointer to the interaction elements requires sufficient attention and support for operating the pointing device which makes it difficult for rough environments.

Example: Microsoft Windows operating system

Most personal computers in use today provide an interface for their users that employs windows, icons, menus, and pointing devices (WIMP). Microsoft introduced its first GUI-capable operating system, Windows 1.0, in 1985. Microsoft's Windows, through repeated release of upgraded versions, has become the world leader in operating system sales. With the rapid rise of Microsoft Windows, WIMP interfaces have become the dominant interface paradigm.

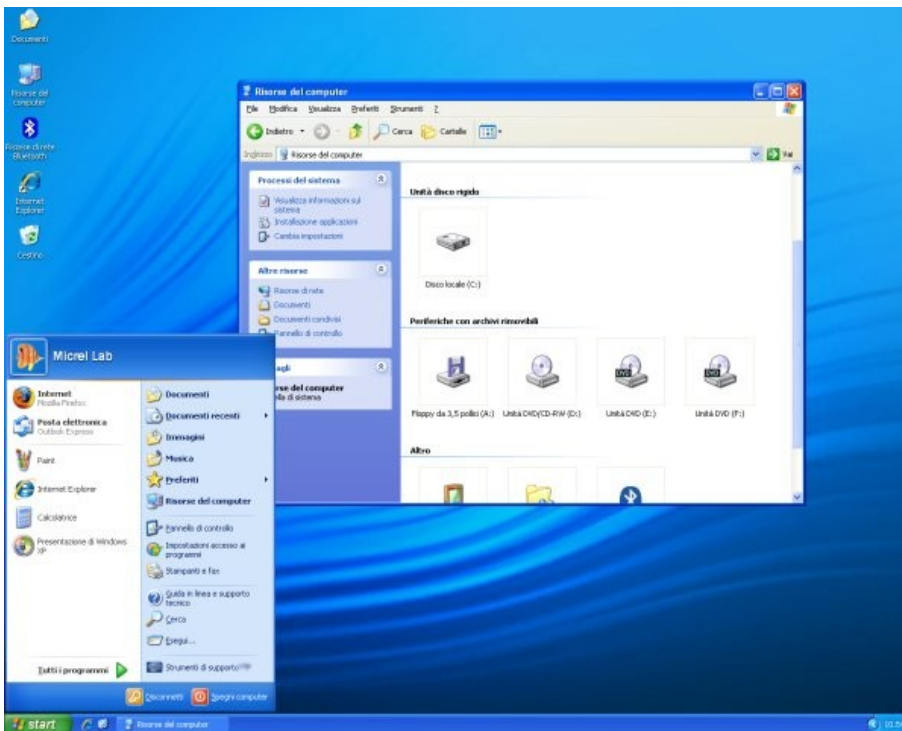


Figure: Windows operating systems are based on WIMP interface

References

Engelbart, D., Friedewald, M., Institute, S. R., & Alliance, B. (1962). Augmenting human intellect: A conceptual framework: Stanford Research Institute Menlo Park, CA.

2.12 One-handed Interaction

One handed interaction is the predominant interaction paradigm for mobile handsets.

For operating a mobile phone in as many situations as possible, it is important to be able to fully access the functions of the phone one-handedly, i.e. while holding and operating the phone in one hand only. Examples for such situations are if users are carrying a bag with the other hand or holding on to a handle in a tram, say.

The typical way of operation is through a combination of buttons and a display, (although pure touchscreen handsets are gaining popularity, see below). For the 'traditional' display and buttons handset (see e.g. the picture below), in general there is a close coupling of haptic input/output and visual output, e.g. when navigating a menu structure through a cursor rocker switch, or when selecting soft keys depending on the decoration on screen. One handed interaction for mobile phones typically includes ways to input text. Especially users that are very familiar with mutlitap or similar input schemes are able to correctly text a message without looking at the display.



The requirement to work one-handedly excludes the usage of a pen, as you need two hands to hold the device and the pen in general. It is however possible to utilize touch screen input (including gestures) with bare fingers, which is of course not as precise as when using a pen and has as a consequence the exclusion of utilizing general handwriting recognition. Typically, for touch screen handsets, the selection by finger is restricted to few icons that may be organized in hierarchies, and sometimes real world metaphors such as simulating inertia while scrolling through lists is used such as with the Apple iPhone to ease operation.

In general, typing text is not as easy with pure touch screen devices, as the tactile feedback and reduced positioning accuracy of bare fingers makes it difficult or impossible to operate without looking at the display. Despite actual market availability, there are alternatives available such as EdgeWrite (see US2004196256) for solving the 'blind' writing issue or the CharacterPump or EnChoi to allow for efficient and easy-to-learn text input with small screens.



Figure: *EnChoi is following the one-handed interaction paradigm, as operations are designed to allow a touchscreen text input with the bare thumb of the same hand that holds the device. Users can slide a letter ribbon up and down, whereby language statistics control for the size of the individual letter fields. By sliding to the left, the letter is entered (or removed again by sliding to the right). The screenshot shows a text search for 'Watchtower' in a music player application, two letters have already been entered. The database is incrementally searched and the list being updated accordingly.*

2.13 Form-based Interaction

A user interface paradigm that has been around for a long time is form based interaction. It basically is a direct adaptation of form filling on paper on computerized systems.

Back in the early days of computing, when we had 'host' or 'main frame' computers, there were 'screen masks' like unfilled forms being sent to remote terminals, then, after operators filled in the required information into the individual fields, they could push a specific send-button on their terminal's keyboard to transmit the filled form back to the host where it was processed. As a result,

a following form was sent to the terminal, so interaction between an operator and the host was done by filling in forms locally and sending back and forth filled or unfilled forms over the communication line.

Although we might think that this kind of interaction is long away, it is still very present with web applications, where users must fill in form fields in their browser, click on a 'submit' or 'send' or 'save' button just to receive another form they have to fill in. Also in commercial software these kind of interaction are still common place.

An advantage of the paradigm is -- besides its easy implementability -- that it is easy to understand for the users (drawing on the strong metaphor of a paper form). On the other hand it is often implemented in a way that makes it difficult for users to make free choices on what to do first, or difficult to operate in presence of partially unknown information.

Form based interaction can be both an interaction paradigm and a 'one-among-others' interaction technique. It can be mixed with other interaction techniques that are more suitable in other areas and form filling is kept for situations (only) for which it is especially well suited.

2.14 Direct Function Control Mapping

With direct function control mapping we understand that for each and every function of the system, there is a dedicated control (switch, dial, lever, indicator, gauge etc.) for it.

A prototypical example for direct function control mapping is the cockpit of an airplane that is full of switches, lamps, instruments and levers. The advantage lies in the fact that each function can directly and instantly be accessed. A (luckily not followed) differing approach would be to group all the functions of a cockpit into a hierarchical menu structure and access these from a graphical user interface. It is obvious that for reasons of safety, the crew must be able to access certain functions very quickly and thus navigating a menu would be too slow. Also, while looking at an instrument such as an artificial horizon, the haptic and tactile feedback of the position of a lever or the wind pressure on the flaps that can be felt is an important information for the pilot.

Also virtually all cars until around 1990 followed the direct function control mapping paradigm, by which time the car industry started to combine functionality in menu structures or context dependent soft keys in order to not overload the dashboard with controls for the ever increasing functionality (possibly mainly for aesthetic reasons). Note that still many often needed functions are directly accessible, such as the volume knob of the radio etc. All safety-relevant controls such as for windscreen wiper, headlights, brake of course will probably always be kept in the direct function control mapping.

In comparably simple systems such as household appliances, the direct function control mapping is also very popular.

As a potential disadvantage, for complex systems such as an airplane the user is required to go through lot of training in order to be able to quickly select the right control, which -- depending on system complexity -- can make it difficult to operate such a system also for an intermittent or first-time user. On the other hand, direct function control mapping is conceptually easy to grasp (given a basic understanding of the functions and proper description of the controls) which makes it -- for

simpler systems -- very suitable for casual users. Also, whereas in direct function control mapping systems, a user has to remember the location of a control in order to operate efficiently, also in menu systems users are frequently not able to spot a function in a complicated menu structure easily. The quality of establishing a good hierarchy and grouping for a menu system is comparable to the quality of a good layout and grouping of controls in direct proximity for related functions.



Figure: the cockpit of an airplane

2.15 Ambient Interaction

Ambient interaction is about communicating information at the periphery of human attention by using "ambient media" such as ambient light, sound, airflow etc. (MIT MediaLab, <http://tangible.media.mit.edu/projects/>). The concept of ambient interaction is based on the idea that humans are very well in monitoring auditory or visual cues in the periphery of their attention. Prominent changes in these ambient cues will be notified by users and may trigger them to have a more explicit interaction based on one of the other interaction paradigms described.

This builds further on the vision of Weiser and Brown (1996) who introduced the notion of "calm technology". The background of this idea is that many devices today behave in ways which makes it difficult for people to ignore their presence in daily life. Calm technology engages both the center and the periphery of our attention, and in fact moves back and forth between the two (Weiser and Brown, 1996).

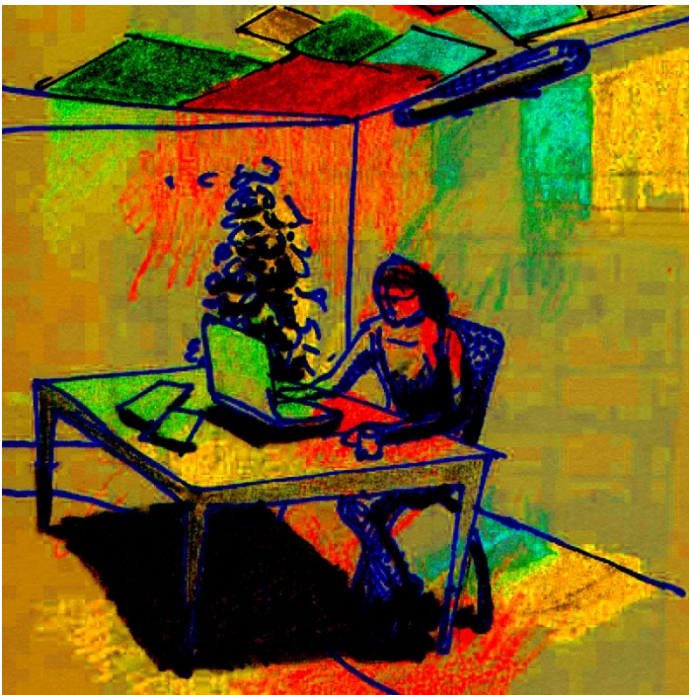
The basic idea of ambient interaction is that the physical environment of a user is used as a vehicle for digital interaction. The user either communicates voluntarily with a device in order to extract detailed information from it or she/he may be discovered by the ambient." This UI may be exploited through the following paradigms and/or a combination of them:

- Augmented Reality
- Tangible User Interfaces (TUIs)
- Object identification

- Visual Object identification: This interaction paradigm allows the user to identify her/his object of interest by simply shooting a photo of the object.
- Blind Object identification: This interaction paradigm allows the user to find her/his object of interest by transmitting to or receiving from the ambient its ID code.

Example: HomeRadio concept

One example developed at Philips Research is the concept of HomeRadio (Eggen, 2003). The idea of this concept is that it addresses the need expressed by families to stay in touch with their home, extending the home experience beyond the boundaries of the physical house. The HomeRadio concept is based on the idea that home activities can be coded by the corresponding utility streams they generate (gas, electricity, water, communication and information). This coded information is broadcast and family members can tune in to this stream. At the receiver's site (e.g. one's office) the coded information is rendered and presented by audio-visual means.



References

Weiser, M. and Brown, J.S. (1996). The Coming Age of Calm Technology. PowerGrid Journal v 1.01. <http://www.teco.edu/lehre/ubiq/ubiq2000-1/calmtechnology.htm>

Eggen, Rozendaal, Schimmel, (2003) HomeRadio: Extending the Home Experiences beyond the Physical Boundaries of the Home. <http://www.crito.uci.edu/noah/HOIT/HOIT%20Papers/HomeRadio.pdf>

2.16 Tangible User Interface

With a Tangible User Interface (Ishii & Ullmer, 1997), a user interacts with digital information through physical environment. Generally tangible interfaces are systems relating to the use of physical artifacts as representations and controls for digital information. A central characteristic of tangible interfaces is the seamless integration of representation and control, with physical objects being both representations of information and as physical controls for directly manipulating their underlying associations. Input and Output devices fall together. There are 4 characteristics concerning representation and control:

1. Physical representations are computationally coupled to underlying digital information.
2. Physical representations embody mechanisms for interactive control.
3. Physical representations are perceptually coupled to actively mediated digital representations. (visual augmentation via projection, sound...)
4. Physical state of tangibles embodies key aspects of the digital state of a system.

TUIs are usually persistent: turn off the electrical power and there is still something meaningful there what can be interpreted.

Tangible user interfaces are often used as input channel in other paradigms. For example, in ambient interaction, Tangible User Interfaces (TUIs) introduce physical, tangible objects that augment the real physical world by coupling digital information to everyday physical objects. The system interprets these objects as part of the interaction language. TUIs become the representatives of the user navigating in the environment and enable the exploitation of digital information directly with his/her hands. Users, manipulating those objects, inspired by their physical affordance, can have a more direct access to functions mapped to different devices. The design of TUI leads to several advantages (Fitzmaurice, 1995)

- it pushes for two handed interaction
- facilitates interaction by allowing interface elements manipulation through physical artifacts
- takes advantage of our spatial reasoning skills
- afford multi-person collaborative use
- externalizes traditional computer representations

Example: Follow-Me Tokens

At Philips Research a tangible interaction concept has been developed for moving activities around in a Connected Home (Van de Sluis, 2001). Physical follow-me tokens were developed which represent individual media activities and could be moved to another room in order to enable end-users to easily relocate activities in the home. The interaction with the tokens was kept very simple. Users only had to pick up a token at the initial location and to put down at the destination location. This was enabled by RFID technology. Each Token contains a transponder and each room has a detection area so that a Token can be detected when it arrives or leaves. This transponder

technology also prevents the user from being bothered with energy management of the tokens.

Besides taking an activity along 'in space' a Token also makes it possible to take an activity along 'in time'. For instance, in the case of a movie, it is possible to pause the movie when a Token is picked up from the vase and to resume it when this Token is put into a vase at a certain location. This feature also makes it possible to use a Token to 'keep' the rest of a movie to be watched later on by just picking up a Token and putting it aside.



Figure: Follow-me tokens for the Connected Home

Example: TANGerINE framework

TANGerINE (TANGible Interactive Natural Environment) is a tangible tabletop environment where users manipulate smart objects in order to perform actions on the contents of a digital media table (Baraldi, 2007). Unlike other approaches, here users are able to interact with the system in three context: the active presentation area (on the surface of the table), the nearby area (around the table) and the external space (a transitional space between tabletops). This is possible thanks to the SMCube (Smart Micrel Cube) a smart wireless object able to recognize user gestures and act as an interface in the three contexts.

Multiple cubes can coexist within the same scenario. In this case the cube acts also as a mean to authenticate users for different activities. Furthermore as users interact naturally with digital elements on interactive surfaces, they are now able to transport data across different contexts just carrying with them a tangible object.

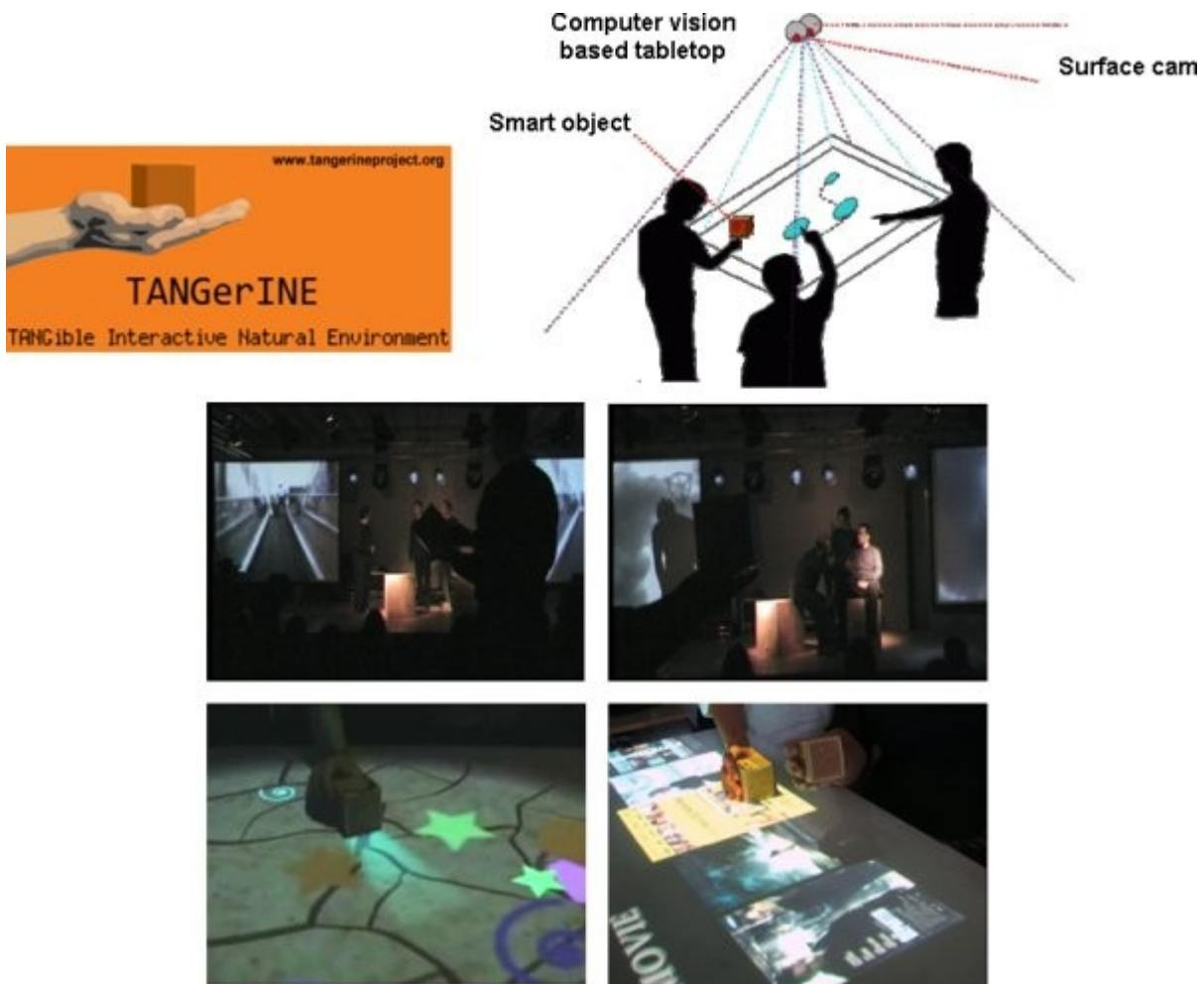


Figure: TANGerINE system

References

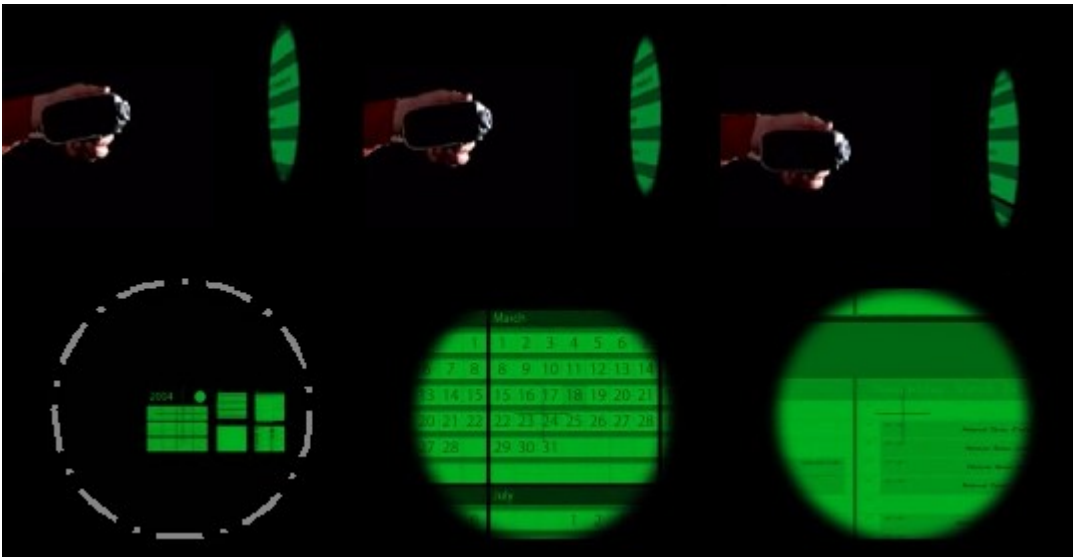
- Ishii, H., & Ullmer, B. (1997). Tangible bits: towards seamless interfaces between people, bits and atoms. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.
- Fitzmaurice G. W., et al. (1995). 'Bricks: laying the foundations for graspable user interfaces'. In /CHI '95: Proceedings of the SIGCHI conference on Human factors in computing systems/, pp. 442-449, New York, NY, USA. ACM Press/Addison-Wesley Publishing
- Van de Sluis, Eggen, Jansen, Kohar (2001), User Interface for an In-Home Environment, INTERACT '01, Tokyo
- Baraldi S., et al. (2007). Introducing tangerine: a tangible interactive natural environment. In Proceedings of the 15th international Conference on Multimedia. Augsburg, Germany.

2.17 Zooming user interface

Zooming user interface (ZUI) was introduced 1993 with a “believe” that navigation in information spaces is best supported by tapping into our natural spatial and geographic ways of thinking (Perlin & Fox, 1993). A ZUI is often a sort of 3D GUI environment, in which users can change the scale of the viewed area in order to see more detail or less. Instead of windows in GUI, ZUI presents information elements directly in an infinite virtual surface, in which the user can either pan around the surface (two dimensions), or zoom into objects of interest (the third dimension of depth).

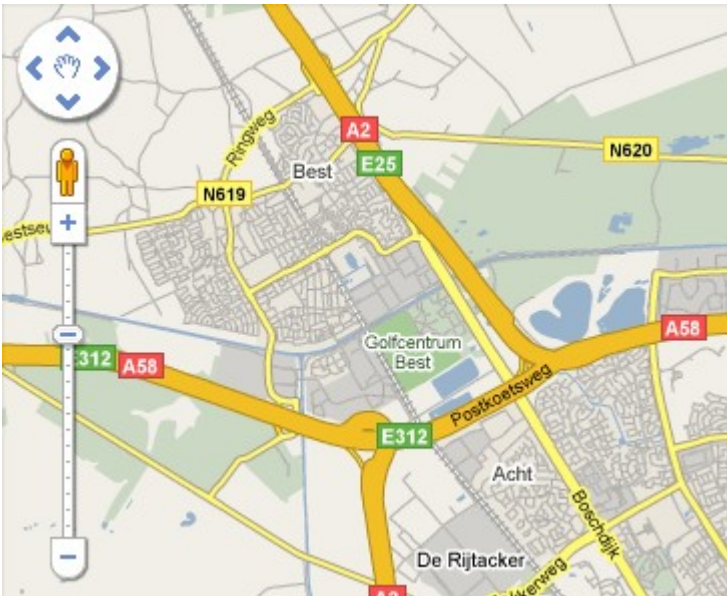
Example: Spotlight Navigation

CONANTE's Spotlight Navigation is a ZUI for handheld projectors that is controlled by direct pointing gestures. The device is operated similar to a flash light, only that it can be used to successively explore a virtual data space rather than just the physical world. Information access is done by panning -- steering the lightbeam in different directions (top row) and zooming -- operating a scrollwheel on the top of the projector with the thumb (bottom row). The virtual information space is infinite and can be arbitrarily enlarged.



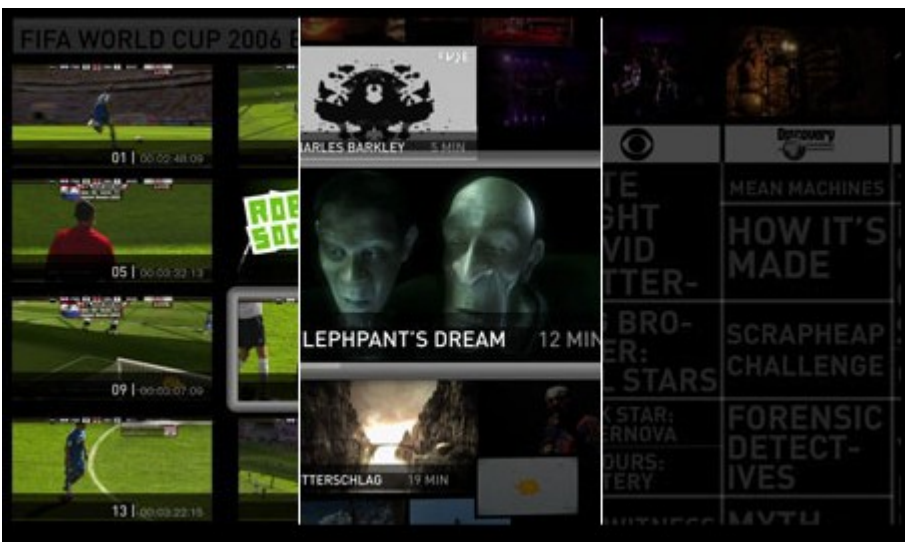
Example: Google Maps

Zoomable world map with integrated search.



Example: OpenTV

OpenTV provides a zoomable user interface (ZUI) that fundamentally changes the way viewers navigate and make viewing choices from the massive amounts of available content, by providing navigation tools that create relevance and match interests.



References

Perlin, K., & Fox, D. (1993, 1993). Pad: An alternative approach to the computer interface. Paper presented at the ACM SIGGRAPH.

2.18 Cultural Computing

Over last few decades, the paradigm of human–computer interaction has evolved from personal computing towards cooperative and social paradigms. The computer is no more centre of interest, nor is it the focus of the attention of the user. It is the benefits and effects on the user that matter. Along this line of evolvement, a new paradigm is proposed to address the cultural determinants (attitudes, norms, values, beliefs, actions, communications and groups, etc.) that have since ancient times, a strong influence on our ontology and epistemology, and the influence of computing on these cultural determinants and vice versa (Hu, Bartneck, Salem, & Rauterberg, 2008; Rauterberg, 2006). This new paradigm is based on Kansei-Mediated Interaction (Nakatsu, Rauterberg, & Salem, 2006). Kansei Mediation is a form of multimedia communication that carries non-verbal, emotional and Kansei information (e.g. unconscious communication). The main research objectives in Kansei-Mediated Interaction are underlying almost unconscious cultural determinants. In this context, cultural computing is more than integrating cultural aspects into interaction. It is about allowing the user to experience an interaction that is closely related to the core aspects of his or her culture (the cultural determinants). As such, it is important to understand one’s cultural determinants and how to render them through the interaction.

Example: Alice's adventure in wonderland

From the narrative of Alice’s Adventures in Wonderland, six stages were chosen, each represents a chapter or part of it. From start to end the user undergoes immersion that consists of real and nature mimicking, virtual and augmented reality in such situations which demands the user to question him/her self and their logic and Western reasoning.



Example: ZENetic computer

ZENetic Computer is a means of cultural translation using scientific methods to represent essential aspects of Japanese culture. Using images—deriving from Buddhism and other Asian concepts, sansui (landscape) paintings, poetry and kimonos— that have not heretofore been the focus of computing, the style of communication developed by Zen schools over hundreds of years is projected into an exotic computing world that users can explore. Through encounters with Zen

koans and haiku, the user is constantly and sharply forced to confirm his or her selfawareness for purposes of the story. There is no one right answer to be found anywhere.



References

- Hu, J., Bartneck, C., Salem, B., & Rauterberg, M. (2008). ALICE's adventures in cultural computing. *International Journal of Arts and Technology*, 1(1), 102-118.
- Nakatsu, R., Rauterberg, M., & Salem, B. (2006). Forms and theories of communication: from multimedia to Kansei mediation. *Multimedia Systems*, 11(3), 304-312.
- Rauterberg, M. (2006). From personal to cultural computing: how to assess a cultural experience. *uDayIV--Information nutzbar machen*, 13-21.

2.19 Distinction of explicit vs. implicit interaction

Often a distinction is being made between explicit and implicit interaction. Explicit interaction refers to situations in which users intentionally interact with a system or environment, whereas implicit interaction refers to cases in which there is interaction between user and system while users are not taking explicit actions to interact with the system or environment. For instance, when a system adapts its behaviour as result of detecting presence or activities of users, this is called implicit interaction.

3 Interaction Technologies

3.1 Introduction

An inventory of interaction techniques and interaction technologies can never be complete, as new techniques emerge all the time in the user interaction research field. Even if we would limit ourselves to a list of techniques known at a specific point in time, the effort would be tremendous if not unfeasible simply due to the huge number of known techniques. In consequence, we try to make a balance and at least briefly cover the main, commonplace, standard and frequently used techniques and additionally present some techniques in more detail that we feel relevant for the project's scenarios (as far as we know them already), or where partners of the consortium have a specific expertise.

As a solution that fits the project's aims, we collect and list these interaction technologies not only in deliverable D4.11, but also maintain the directory active as a 'living document' on the project's wiki, so that whenever we see that an interaction technique not covered initially should be considered for a scenario, it can be added there and thus can easily be referred to and communicated to the developers of the other work packages.

3.2 Modality perspective

If we consider the interaction techniques or interaction technologies, we can take a usability/human-centric or a technical/device viewpoint. We have decided for the first one and will group the interaction technologies in the following by their dominating modality. Although elaborate theories on modalities in Human Computer Interaction (HCI) exist (e.g. Bernsen93), we constrain ourselves to a categorization to a general/basic understanding of modalities or human senses: visual, auditory, haptic (both touch and proprioception), taste, smell, thermoception (temperature), nociception (pain), and equilibrioception (balance). We thereby consider that the computerized system also has 'senses' just like the human, although this is obviously strictly speaking not the case. In doing so, we believe this helps the reader in categorizing the technologies into meaningful chunks of information.

Inside each modality, we further differentiate between input, output and input/output to ease the access for the reader further. "Input" stands for an input technology that is, a technology that is used primarily for input from the user to the computerized system, "output" is for technologies concerned with giving information from the system to the user, and "input/output" is chosen for technologies that rely on a mixture of input and output. Often, a technology relies to some extent also a bit on another sense, or as an input technology just a little bit on output as well, or vice-versa. We neglect this kind of relation in the categorization and may or may not mention it in the description.

References

Niels Ole Bernsen (1993) "Modality Theory: Supporting Multimodal Interface Design", in: ERCIM

3.3 Visual

The visual modality is mainly used to output information to the user. However, with the availability of visual sensors in the form of cameras included in virtually all mobile devices, and with the increased processing power, visual input technologies have gained more interest in recent years.

3.3.1 Input

3.3.1.1 Video scene analysis

Activity detection is a generic term which is highly dependent on the application. A variety of approaches exist with no method giving superior results. The term activity can have different meaning depending on the level of details with which a scene is observed. At a long distance, activity is simply the number of people and their motion. At closer ranges, body posture conveys information about activity. At an even closer look, facial expression and gaze direction become relevant. It is thus important to have a clear view of the context in which activity should be detected.

In the context of a shop environment, activity can be any video-captured phenomenon that provides useful information about the customer. Typical useful information is arranged from further away to close up: number of people, occupation of areas, trajectory of each subject, body posture and gestures, gender and age of each person, facial expression and gaze direction. 1) In the field of human detection by video camera, many methods focus on pedestrian detection. In the approach proposed by Dalal and Triggs (2005), histograms of gradients provide excellent performance. Unfortunately, such cost comes at relatively high computational cost. With current computing technology, it is possible to process 320x240 pixel images at 1 frame per second using a sparse scanning methodology that evaluates roughly 800 detection windows per image. 2) In the field of human tracking by video camera, it is still a challenging to track multiple objects efficiently and robustly in complex environments. In shop environments, the difficulties come from cluttered background, complex motion dynamics, occlusion, lighting variations, and so on. 3) In the field of posture estimation by video camera, the posture of a person conveys important information about the internal state of the subject. Estimation of human body pose is a complex problem, given the high variability in appearance of people and the large amount of freedom of configuration of a body with many degrees of freedom. A recent overview is presented by Jackson et al. (2008). 4) In the field of gender / age estimation by video camera, a large number of studies have investigated gender/age classification by human faces (Moghaddam and Yang, 2002). Automatic customer categorisation into age groups and gender is primarily interesting to target advertisements.

While human detection is improving rapidly and will soon achieve the same robustness of face detection, human pose estimation is still an open problem, and no clear technique is prevailing. The main limitations are: the prohibitive amount of training data needed; the assumption of background with limited or no clutter; the estimation speed often far from real time. Age and gender estimation are seldom addressed in real world scenario, with low resolution images and poor of dynamically changing lighting conditions. In order to effectively adopt these methods in shop environment vast improvements on these points must be carried on. Gaze estimation at a distance remains as difficult

as attractive. In situations in which active infra-red and cameras with a strong zoom cannot be adopted, estimating the pose of the face could be an attractive, even if less accurate, solution.

Over the last ten years, the dominant paradigm for almost all robust tracking systems has been to use some form of Bayesian estimation (Piater, 2002). With this approach, targets are detected and tracked in 2D (as opposed to 3D) using simple pixel level operations based on such features as color (Wren, 2003), motion (Chomat, 1999), texture, or adaptive background subtraction (Stauffer, 1999), followed by grouping into moment based blobs to estimate the first and second moments of connected blobs of pixels (Crowley, 1997). Such 2D tracking is attractive because it can easily be implemented to run on-line, at video rate, using embedded computing platforms with limited computer power (typically less than 1 Ghz processors).

When used with simple pixel level colour detection, such tracking makes possible face tracking for camera control and face recognition (Schwerdt 2000) as well as hand and finger detection (Martin 1997) for gesture interpretation. When used with motion detection and adaptive background subtraction, blob tracking makes possible a large variety of video surveillance applications for security and commercial services (Piater 2002). When used with pixel level Bayesian estimation, such methods are easily extended to using local appearance for detecting certain classes of targets.

When people cross in front of a camera, which is likely to occur in an in-store shop environment, 2D tracking from the problem of "split" and "merge" of targets. Such problems are often addressed by learning probabilistic methods for detecting 2D "Layers" (Darrell 1991). Nonetheless, the most robust and effective method for separating overlapping blobs is to provide depth information from multiple cameras (Ferrer-Biosca 2006). With this approach, estimated 3D target blobs are used to predict the 2D position and size for each camera. The 2D detection of blobs within a detected region is then combined with camera position and orientation information to update an estimate of 3D blob position.

The problem of coalescing targets remains a difficult task when tracking multiple targets simultaneously, especially if these targets are similar or even identical in appearance. While it is often possible to assign independent tracker labels to individual targets, in 2D systems, targets can coalesce, resulting in a loss of track.

References

- Dalal N., and Triggs, B. (2005). Histogram of oriented gradients for human detection. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 886–893, 2005.
- Jackson, J., Yezzi, A.J., and Soatto, S. (2008). Dynamic shape and appearance modeling via moving and deforming layers. IJCV 2008.
- Moghaddam, B., and Yang, M. (2002). Learning gender with support faces. IEEE Trans. Pattern Analysis and Machine Intelligence, 24(5):707-711, May 2002.
- Stauffer, C., and Grimson, W.E.L. (1999), Adaptive background mixture models for real-time tracking, in: IEEE conf on Computer Vision and Pattern Recognition, CVPR 99, Fort Collins, Colorado pp. 246–252. (Jun.1999).
- Wren, C., Azabajejani, A., Darrell, T., and Pentland, A. (1997). Pfnder: Real-time tracking of the human body. IEEE Transactions on Pattern Analysis and Machine Intelligence 19 780–785.

Chomat, O., and Crowley, J.L. (1999). Probabilistic Recognition of Activity using Local Appearance, IEEE Conference on Computer Vision and Pattern Recognition, CVPR 99, Fort Collins, June 1999.

Martin, J., and Crowley, J.L. (1997). An Appearance-Based Approach to Gesture-Recognition, 9th International Conference on Image Analysis and Processing, Florence, Italy, September 1997

Crowley, J.L., and Berard, F. (1997), Multi-Modal Tracking of Faces for Video Communications, IEEE Conference on Computer Vision and Pattern Recognition, CVPR '97, St. Juan, Puerto Rico, June 1997.

Piater, J., and Crowley, J.L. (2002). Event-based Activity Analysis in Live Video using a Generic Object Tracker, Performance Evaluation for Tracking and Surveillance, PETS-2002, Copenhagen, June 2002.

Chomat, O., and Crowley, J.L. (2000), A Probabilistic Sensor for the Perception of Activities, 4th IEEE International Conference on Automatic Face and Gesture Recognition", FG00, Grenoble, France, March 2000.

Schwerdt, K., and Crowley, J.L. (2000). Robust Face Tracking using Color, 4th IEEE International Conference on Automatic Face and Gesture Recognition", FG00, Grenoble, France, March 2000.

Darrell, T. and Pentland, A. P. (1995). Cooperative robust estimation using layers of support. IEEE Trans. on Pattern Analysis and Machine Intelligence, 17(5):474--48.

A. Ferrer-Biosca and A. Lux , (2006) A Visual Service for Distributed Environments: a Bayesian 3D Person Tracker, INRIA Internal report, June 2006.

3.3.1.2 Visual object categorization

Visual object categorization aims to detect objects in images and to determine the object's categories (e.g. cars or humans). This is in contrast to the recognition of specific, individual objects (e.g. my car or Barack Obama). Humans are generally wide better in detecting an object's category than in specific recognition, whereas categorization turns out to be much harder for today's computers and algorithms than object recognition. The main problem is the definition of the elusive concept of "visual category", which usually has to gather instances with large intra-class variability and to discriminate classes with marginal inter-class differences.

Obviously, in order to identify and learn a visual category, raw images must be processed and features extracted to reduce the amount of data while (hopefully) maintain the discriminative information. This feature set may use many facets of object in images: color, texture, discontinues as edges or corners, homogeneous regions, and spatial relationship between some or all of these local features. However, every simple feature that can be measured in an image will change when certain parameters of the image capturing process and/or illumination and spectral sensitivity of the sensor are changed. Hence, features that are invariants against these sources of distortions play a central role in the successful solution of the object categorization problem.

A representation based on invariant local features typically extracts feature vectors by applying some sort of detector of salient, distinguish points in the image. One of the older but still used one is the Harris corner detector (Harris 88), which in its original formulation was only rotational invariant but has recently been extended to cope with scale invariance (Mikolajczyk 01) and affine invariance

(Mikolajczyk 02). Another major category of detectors is comprised of “blob” detectors: they fire on almost circular homogeneous regions of the image. Recently these detectors have been proposed in conjunction with very effective feature descriptors; hence they are among the most used ones. Examples include the Laplacian and the Determinant of the Hessian (Lindeberg 98), the DoG detector (Lowe 04) and the FastHessian detector (Bay 06). These detectors are designed to be scale-invariant and their repeatability falls dramatically between images of a same object taken with an angle greater than 25-30° degrees. Similarly to what has happened for the Harris detector, an extension to obtain affine invariance has been proposed for the Hessian based detectors (Mikolajczyk 02). Finally, detectors may look for distinguish regions, regardless of their shape. These detectors are typically referred to as region detectors and have been proposed in the form of Intensity based regions (IBR) (Tuytelaars 04) and Maximally Stable Extremal Regions (MSER) (Matas 02). They are usually affine invariant by design. In a recent survey (Mikolajczyk 05), no detector emerged as the “one size fits all” choice, but in general MSER and Hessian-Affine may be expected to obtain the higher repeatability. The MSER detector usually finds less keypoints than Hessian- or Hessian-affine, but is also by far more computationally efficient.

When salient regions or points with their support region have been detected, a representation able to easily allow for searching correspondences among features must be selected. The algorithm that produces an (invariant) feature vector from a detected region or a patch surrounding an interest point is usually referred to as descriptor. Simple descriptors may be raw gray-level patches. These redundant feature vectors may be compared by Normalized Cross Correlation (NCC). They are typically used with geometric and photometric affine invariant detectors, normalizing the detected region to a canonical patch. Moments, and especially invariant moments (Hu, 62), are another simple and popular choice, extended nowadays to affine and photometric invariance (Van Gool, 96). SIFT (Lowe 04) and SURF (Bay 06) are recent, scale invariant proposals, based on histogram of gradient orientations and responses to the Haar wavelets, respectively. They have gained momentum in the last years and especially the SIFT success has paved the way to the spread of local features in almost every field of Computer Vision.

As far as categorization is concerned, descriptors are used in almost every representation of visual categories. The role and importance of visual descriptors, however, has faced a shift in importance. People have tried with success to perform categorization computing descriptors at every pixel in the images or at a rather dense grid (Bosch 06). Both keypoint based and grid sampling approaches have pros and cons and performances are mainly related to training and test datasets. For prominent objects and little background clutter, saliency detectors are recommended as an efficient means to reduce the amount of data to be processed. Grid based methods may be preferable when information about homogeneous regions is essential and when the object appear at smaller scales in cluttered images (Pinz 06).

In order to model a visual category, keypoint descriptions must be grouped to form a category model. In the bag of visual words proposal, no geometric model is actually used. An image is represented by a vector that comprises information on the frequency of a given visual word in an image. Nothing is stored and used about words spatial location. A visual word is the center of a cluster of similar local features collected across training images, as it results from an unsupervised clustering step (i.e. k-means or agglomerative clustering) performed on all training local features. The set of all the visual words of a category is called a “codebook” and the vector describing a test

or training image simply stores the frequencies of such codebook entries in the given image. The size (granularity) of this vocabulary is an important parameter that may discriminate between a system able to perform object recognition and object categorization. From this feature vector representation, a classifier can be trained with positive and negative examples of a class, and then used in the on-line testing stage. The various approaches of this kind differ mainly in the types of keypoints and descriptors used and in the learning algorithms that are used to obtain classifiers (typical choices are SVMs and AdaBoost) (Csurka 04) (Zhang 05). These proposals are largely unaffected by position and orientation of object in the image, especially if affine invariant feature are used. Furthermore, image description has a fixed length, defined by the codebook cardinality, irrespective of the number of detected keypoints. When the problem is just to categorize an image, they present the current state of the art. If object localization within the image is required in addition, their performances depend on the nature of training and test images, but are generally very poor.

Part based representations try to overcome these disadvantages, using a model that captures local saliency, but also represents spatial relationship between parts. Simpler forms of spatial relationship are obtained using ROIs and/or sliding windows, and a bag of visual word representation of each ROI (see (Lampert 08) for a recent, successful modification of this approach); covering the object with tiles, each of which carries its own bag of visual words (Fergus 05); or covering the whole image with tiles, and computing Histogram of Gradients (HOGs) in every position (Dalai 05). The constellation model (Fergus 03) is probably the most popular part-based representation today. Object are modeled as random constellations of parts, explicitly representing the mutual relationship between parts, as if they were connected by springs. A major limitation of this proposal is that only a limited number of parts may be used and learned, otherwise the problem of learning them become computationally infeasible, since the model is actually a fully connected graph. A recent alternative (Leibe 05), advocates the use of an Implicit Shape Model, where the shape of an object is defined implicitly by the position with respect to the object center that every visual word stores about its concurrencies in the training images. This simplifies the learning stage, since it introduces independence between learning of visual words spatial clues, and allows for theoretically any shape to be represented.

Finally, given the recent advances and results in object categorization based on local features, it may be of interest in many SOFIA use cases, where it may ease and add intuitiveness to the interaction between the user and the system, e.g. in the automatic identification of a broken system/component that requires or has been subject of maintenance operations

References

- M. K. Hu, "Visual pattern recognition by moment invariants," *IRE Trans. Information Theory*, vol. IT-8, pp. 179–187, 1962.
- C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, pages 147–151, 1988.
- L. Van Gool, T. Moons, and D. Ungureanu, "Affine/photometric invariants for planar intensity patterns," in *Proc. ECCV*, pp. 642–651, 1996.
- T. Lindeberg, "Feature detection with automatic scale selection," *International Journal of Computer*

Vision, vol. 30, no. 2, pp. 77–116, 1998.

K. Mikolajczyk and C. Schmid, “Indexing based on scale invariant interest points,” in Proc. ICCV, pp. 525–531, 2001.

J. Matas, O. Chum, M. Urban, and T. Pajdla, “Robust wide baseline stereo from maximally stable extremal regions,” in Proc. 13th BMVC, pp. 384–393, 2002.

K. Mikolajczyk and C. Schmid, “An affine invariant interest point detector,” in ECCV (1), pp. 128–142, 2002.

R. Fergus, P. Perona, and A. Zisserman, “Object class recognition by unsupervised scale-invariant learning,” in In Proc. IEEE Conf. Computer Vision and Pattern Recognition, CVPR, 2003.

G. Csurka, C. Bray, C. Dance, and L. Fan, “Visual categorization with bags of keypoints,” in ECCV Workshop on Statistical Learning in Computer Vision, pp. 1–22, 2004.

D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” International Journal of Computer Vision, no. 2, pp. 91–110, 2004.

T. Tuytelaars and L. Van Gool, “Matching Widely Separated Views Based on Affine Invariant Regions,” International Journal of Computer Vision, vol. 59, 2004, pp. 61-85.

R. Fergus, P. Perona, and A. Zisserman, “Learning object categories from Google’s image search,” in Proc. ICCV, 2005.

W. Zhang, B. Yu, G. J. Zelinsky, and D. Samaras, “Object class recognition using multiple layer boosting with heterogeneous features,” in Proc. CVPR, pp. 323–330, 2005.

K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, “A comparison of affine region detectors,” International Journal of Computer Vision, vol. 65, no. 1/2, pp. 43–72, 2005.

Dalal N., and Triggs, B. (2005). Histogram of oriented gradients for human detection. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 886–893, 2005.

B. Leibe, A. Leonardis, B. Schiele, Robust object detection by interleaving categorization and segmentation, in: International Journal of Computer Vision, 2005.

H. Bay, T. Tuytelaars, and L. Van Gool. SURF: Speeded Up Robust Features. In Ninth European Conference on Computer Vision, 2006.

A. Bosch, A. Zisserman, and X. Munoz, “Scene classification via pLSA,” in Proc. ECCV, pp. 517–530, 2006.

E. Nowak, F. Jurie, and B. Triggs, “Sampling strategies for bag-of-features image classification,” in Proc. ECCV, pp. 490–503, 2006.

A. Pinz, Object Categorization, Foundations and Trends® in Computer Graphics and Vision, 1(4), pp. 255-353, 2006.

Christoph H. Lampert, Matthew B. Blaschko, Thomas Hofmann: "Beyond Sliding Windows: Object Localization by Efficient Subwindow Search", IEEE Computer Vision and Pattern Recognition (CVPR), Anchorage, AK, 2008

3.3.1.3 Facial expression recognition

Damasio has argued that the facial expressions of emotion (as well as some other physiological expressions) precede feelings of emotion (Damasio, 1999). In this view, the facial expression of emotion is fundamental to emotional feelings and reactions, presenting emotional reactions in an easily readable external form. Our problem is to provide the technology to read such expressions.

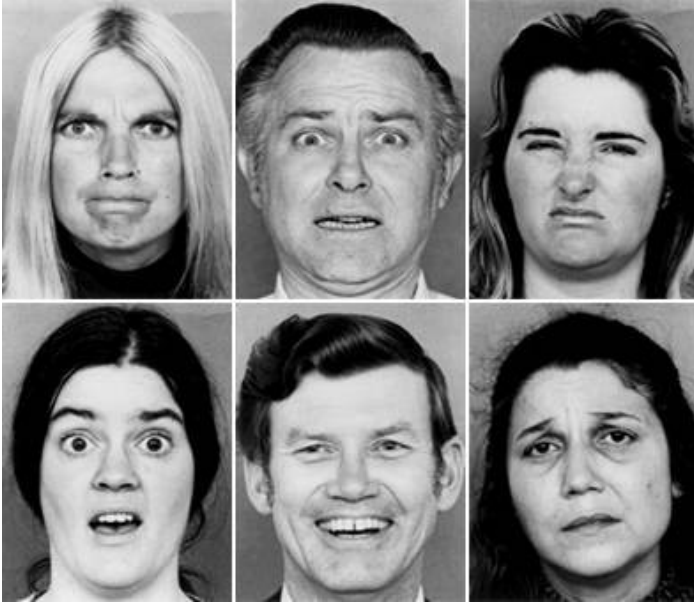


Figure 1. Photographs of facial expression of basic emotions used in cross-cultural studies done by Ekman: anger, fear, disgust, surprise, happy and sad.

Machine perception of the emotional state of humans is a notoriously hard problem that is rapidly gaining attention in all areas of machine perception. Driven partly by the appreciation that computer interfaces are socially insensitive (Reeves, 1996), a movement is growing to develop a science of affective computing (Picard, 1997). The area first gained attention in the mid-90's as many researchers attempted to extend the "Eigen-Faces" approach (Turk, 1991) to recognition of Ekman's Facial Action Codes (FAC) (Ekman, 1977) in face expressions (see also Figure 1). Such efforts proved frustrating, in part because the emotions in face expressions give rise to individual specific visual appearances. Nevertheless, Ekman's work on FAC has led many computer vision researchers to formulate the problem as that of assigning facial image sequences to one of seven fundamental classes (Essa 1997). Work on perceiving emotions from audio has met similar frustrations (Aubergé, 2003), in part for many of the same reasons.

Automatic facial expression analysis generally works along three subsequent steps:

1) Face acquisition is a pre-processing stage to detect and locate the face region in input images or sequences. The real-time face detection scheme proposed by Viola and Jones (2001) is arguably the most commonly employed face detector. To handle large head motion in video sequences, head tracking and pose estimation needs to be adopted. 2) Facial feature extraction and representation tries to derive an effective facial representation from the original face images. Two types of features are usually considered: a. geometric features deal with the shape and locations of facial components (including mouth, eyes, brows, and nose) (Pantic and Bartlett 2007); These features require accurate

and reliable facial feature detection and tracking, which is difficult to accommodate in real-world unconstrained scenarios; b. appearance features represent the appearance changes (skin texture) of the face (including wrinkles, bulges and furrows). 3) Facial expression recognition is concerned with the classification of the possible emotional expressions from the extracted facial features. Depending on whether or not the temporal information is used, the recognition approaches are generally divided as image-based or sequence-based.

Previous work on facial expression recognition has been carried out on expression data that were collected by asking subjects to deliberately pose facial expressions (for instance, the well-known Cohn-Kanade database containing videos of subjects that pose one of the six basic emotions: anger, fear, disgust, surprise, happy and sad, as shown in Figure 1). However, spontaneous facial expressions induced in natural real-life environments are more subtle and fleeting, such as tightening of the lips in anger or lowering the lip corners in sadness. More recent studies try to bring the field forward by focussing on spontaneous and natural expressions, however those studies are typically limited to one or two emotions.

Previous work is also based on the classification of a single image of a frontal face at a time. Recent studies also incorporate information from time instances by a sequence-based approach. The latter is especially required when moving towards spontaneous and natural expressions, since these are typically much more subtly expressed than the posed expressions. In an in-store shop environment, the expressiveness of emotions is expected to be moderate to low, not recorded in frontal view and by limited resolution cameras. Frontal faces are not always available due to natural head movements. The recognition of subtle expressions will be crucial, but difficult. Using multiple modalities in combination for emotion recognition is expected to improve performance.

References

- Reeves, B., and Nass, C. (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*, Cambridge University Press, New York, 1996
- Picard, R.W. (1997), *Affective Computing*, MIT Press, Cambridge, MA, 1997.
- Ekman P., and Friesen, W.V., (1977), *Facial Action Coding System*, Consulting Psych. Press, 1977.
- Turk, M., and Pentland, A. (1991). Eigenfaces for Recognition, *Journal of Cognitive Neuroscience*, vol. 3, pp. 71-86, 1991.
- Essa, I.A., and Pentland, A.P. (1997). Coding, Analysis, Interpretation, and Recognition of Facial Expressions, *IEEE Trans. PAMI*, vol. 19, pp. 757-763, July 1997.
- Aubergé, V., and Cathiard, M. (2003). Can we hear the prosody of a smile? *Speech Communication*, vol. 40, pp. 87-97, 2003.
- Pantic, M. and Bartlett, M.S. (2007). Machine analysis of facial expressions. In K. Kurihara (Eds.), *Face Recognition*, pages 377-416. *Advanced Robotics Systems*, Vienna, Austria, 2007.
- Shan, C. (2007). *Inferring Facial and Body Language*, PhD Thesis, University of London, London, UK, October, 2007
- Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *IEEE International Conference on Computer Vision and Pattern Recognition*, pages 511-518, 2001.

3.3.1.4 VideoBasedGestureRecognition

The methods described in the section video scene analysis can be utilized to detect gestures in interaction. From an interaction point of view, the detection of pointing gestures is important (Weller93, Sá&al2001) to identify objects in the real world or information rendered on-screen. Also the detection of gestures like head nodding or waving hands is utilized in HCI (Daugman1997, Morency&al2007).

References

Wellner, P., Interacting with paper on the DigitalDesk. Communications of the ACM, 1993. 36(7): pp. 87 - 96.

Sá, V.; Malerczyk, C.; Schnaider, M.: Vision Based Interaction within a Multimodal Framework 10th Portuguese Computer Graphics Meeting, ISCTE, Lisbon, 2001.

Daugman, J. 1997. Face and Gesture Recognition: Overview. IEEE Trans. Pattern Anal. Mach. Intell. 19, 7 (Jul. 1997), 675-676.

Morency, L.P. and Sidner, C. and Lee, C. and Darrell, T.J. (2007) Head gestures for perceptual interfaces: The role of context in improving recognition, AI, vol. 171, 2007, 8-9 June, pp. 568-585.

3.3.2 Output

3.3.2.1 TextDisplay

One of the simplest ways to put information from the system to the user is via linguistic textual expression. For this, text displays have a long history, basically from the early days of computing when teletype writers have been hooked up to the first computers, then more and more teletypes having been replaced by CRT-Terminals. Also with the advent of graphical terminals, displaying a textual message is the dominant form of giving information to a user. And still with graphical user interfaces, most of the information is still encoded in text form.

Also in embedded electronics and consumer electronics, text displays in the form of one- or two-line LCD displays are very widespread and in use in a very broad range of devices, from printers and copiers over home stereos and DVD-players to coffee machines. An advantage here over status LEDs is that more different messages can be transported to the user. Also, ideally, the messages are directly understandable by the user without the need to consult a manual. On the other hand, as with all linguistic information, text messages are language dependent and manufacturers have to make sure that messages are available in a language users are fluent in.

3.3.2.2 Menus

Menus is an interaction technique which is common to windowing and non-windowing systems. Its basic implementation consists of a presentation of operations or services in the form of vertical and/or horizontal lists that can be performed from the system at any time. For this purposes the names provided in the menus need be meaningful and informative. Selection of menu voices usually involves other forms of user interaction such as pointing, clicking or touching. Several layers of cascading menus need be designed when a list should contain too may items: to do so an item

selection can open up another menu. Many research efforts have been dedicated in designing menu systems, especially hierarchical ones that reduce traversal time, and associated techniques to evaluate their usability and performances. Specialized techniques have been proposed recently for mobile phone displays which are usually small allowing only few menu items to be displayed (Amant, 2004). A new menu system, the Jumping Menu, has been introduced by (Ahlstrom 2006) which warps the screen cursor to the right into open sub-menu levels when a mouse click is detected inside a parent item. The Jumping Menu results in facilitating interaction and is a promising alternative to conventional pull-down menus. In Jumping menu systems long and hard to perform horizontal movements are avoided and only vertical cursor movements are needed to reach an arbitrary menu item. In order to choose the proper menu interaction the time needed to select an item is the performance index.

References

Dix, et al. Human computer interaction Pearson Education, 2004

St. Amant, R. Horton, T. Ritter, "Model-based evaluation of cell phone menu interaction". Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '04) (Vienna, Austria)

D. Ahlstrom, R. Alexandrowicz, M. Hitz, "Improving Menu Interaction: a Comparison of Standard, Force Enhanced and Jumping Menus" (CHI '06) (Montreal, Canada)

3.3.2.3 PieMenus

Pie menus are a special form of menus where the individual items are grouped in a circle rather than in linear (vertically or horizontally). Usually they are used as a context menu, while operating with a pen or mouse, and the menu ring is displayed around the pen tip or mouse cursor. It is claimed that pie menus were first proposed in 1969 (Wiseman&al1969, Kurtenbach2004). Seemingly independently, they have been reinvented in 1986 by Callahan and Hopkin (Callahan&al1988).

As a specific advantage over linear menus, all items are equally distant from the tip or cursor, allowing an equally fast access to all items. Also, as the direction is sufficient in choosing, the right item can be chosen without visual control, which may be another reason for a 15% speed increase and improved selection accuracy. Pie menus can be operated like traditional menus, i. e. by clicking, but there are also variants where crossing a radial border selects the item (e.g. GuimbretièreWinograd2000). Also, like linear menus, pie menus can be hierarchical.

Pie menus received special attention with developers of large interaction areas (such as interactive whiteboards, GuimbretièreWinograd2000) and with graphics programs (alias Sketchbook, now autodesk sketchbook, in the 'marking menus' variety, Kurtenbach93).

References

Wiseman, N. E., Lemke, H. U., Hiles, J. O. (1969) PIXIE: A New Approach to Graphical Man-machine Communication. Proceedings of 1969 CAD Conference Southhampton, 463, IEEE Conference Publication 51.

Gordon Kurtenbach, 2004, "Notes on the History of Radial menus, Pie menus and Marking menus",

unpublished article, <http://www.dgp.toronto.edu/~gordo/papers/Notes%20on%20History%20of%20Radial%20Menus.pdf>

Callahan, J., Hopkins, D., Weiser, M., Shneiderman, B. (1988) An empirical comparison of pie vs. linear menus. Proceedings of the CHI '88, 95-100, New York: ACM.

François Guimbretière, Terry Winograd (2000) FlowMenu: combining command, text, and data entry, Proceedings of the 13th annual ACM symposium on User interface software and technology, p.213-216, November 06-08, 2000, San Diego, California, United States

Kurtenbach, G. (1993). The design and evaluation of marking menus. PhD Thesis thesis. University of Toronto.

3.3.2.4 2D and 3D Graphics

2D – 3D graphics are the interaction technologies which allow to define the geometry of any kind of object in computer generated scenes. With the so called "through-the-window" modeling the user interacts with the software by means of 2D input devices (mouse, graphics tablet,...). Immersive Virtual Environments usually extend the 3D nature of the scenes of the virtual world to the interaction techniques [O'Toole2000]. Most immersive systems give a 3D virtual representation to menus and dialog boxes, too. Hybrid 2D/3D user interface for immersive object modeling are largely exploited [Coninx1997]. Those interactions that benefit most from 3D graphics exploit the 3D nature of the immersive system, such as navigation to explore the design space. 2D interaction techniques, such as X-Windows based menus and dialog boxes, are integrated in the 3D immersive modeling system. In this way the designer can make an appeal on his skills in using traditional 2D GUIs (Graphical User Interfaces) to interact with these widgets.

References

O'Toole B E, 2000, "The Integration of ArcView/3D Analyst and 3-dimensional visualization technologies for interactive visualization of urban environments",

K. Coninx, F. Van Reeth, E. Flerackers, 1997 "A Hybrid 2D / 3D User Interface for Immersive Object Modeling", Proceedings of the 1997 Conference on Computer Graphics International

3.3.2.5 Augmentation

Augmentation is the fundamental (output-related) interaction technique of AugmentedReality, in where virtual information is superimposed on the real world.

Albeit the term Augmented Reality being introduced later, the seminal work of Sutherland on head mounted displays (Sutherland1968) first introduced the idea (and a practical prototype) to augment the real world with virtual, computer-generated images.

Augmentation can be achieved in that a video stream captured by a camera and displayed on a screen or head mounted display is in part replaced by computer generated images, such that for the viewer, a mixture of the (captured) real world scenery and the virtual objects can be seen. Another way is to project the virtual information onto the real world by use of digital projectors (BimberRaskar2005, RappWeber2005). For both approaches, it is important to know the position and orientation of real world objects such that the computer generated image can be superimposed

on the right spot. Also the camera position in the case of camera-based augmentation or the projector position in the case of projection based augmentation must be known (see Camera Pose Estimation for Augmented Reality).

References

Ivan E. Sutherland, A head-mounted three dimensional display, Proceedings of the December 9-11, 1968, fall joint computer conference, part I, December 09-11, 1968, San Francisco, California

Bimber, O. and Raskar, R. (2005) *Spatial Augmented Reality: Merging Real and Virtual Worlds*. A K Peters LTD (publisher), ISBN: 1-56881-230-2, July 2005

Rapp, S. and Weber, I. (2005) LumEnActive: A novel presentation tool for interactive installations, In: 6th International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST 2005), November 8–11, Pisa, Italy.

3.3.2.6 *SituatedDisplays*

The idea of situated displays is generally attributed to Fitzmaurice in 1993. He coupled a small handheld display (a 4 inch color monitor) with a 6DOF tracking system. Depending on the posture (place and orientation) of the screen, users can see different views on a virtual data space. The display acts as a window or porthole to the virtual world, and the content is updated in real time so that the user has the impression of seeing the virtual data world through the display's screen.

Depending on the place, there can be different information visible, for example when moving the device over a real, physical map of Canada, weather information for the relevant spot is displayed.

Some years later, a mechanical tracking system was coupled with a larger, high resolution monitor by art+com to inspect a virtual car in original scale and select different options of the car.

CONANTE's LumEnActive utilizes projection for spatially aware, situated displays. The light cone of a digital projector can be steered by the mirror of a computer controlled reflection unit to serve as display on any suitable surface of a room. Moving can be done for example by a mouse or any other pointing technique. Similar to Fitzmaurice's work, the content that is displayed is directly depending on the direction, allowing an intuitive exploration of the data space by simply panning around the displayed viewport (see also Spotlight Navigation in the section on ZUI).

References

Fitzmaurice, G.W. (1993). Situated Information Spaces and Spatially Aware Palmtop Computers. *Communications of the ACM*, 36(7), 38-49.

Art+Com virtual vehicle (1997)

http://www.artcom.de/images/stories/2_pro_virtfahrzeug/virt_fahrzeug_e.pdf

Stefan Rapp, Irene Weber (2005) LumEnActive: A novel presentation tool for interactive installations In: 6th International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST 2005), November 8–11, Pisa, Italy.

3.3.3 Input/Output

3.3.3.1 Direct Manipulation

Direct manipulation is a term coined at the beginning of the 80's by Ben Shneiderman (Shneiderman, 1983) to describe a new HCI style that involves continuous representation of objects of interest, and rapid, reversible, incremental actions and feedback. The intention is to allow a user to directly manipulate objects presented to them. In particular, the following properties characterize direct manipulation interfaces:

1. Continuous representation of the object of interest
2. Physical actions or labelled button presses instead of complex syntax
3. Rapid incremental reversible operations whose impact on the object of interest is immediately visible.

At the heart of this development is the promotion of graphic and manual forms of interaction over and above more abstract and linguistic ones, with the aim to reduce the user cognitive load. Studies shown how direct manipulation can considerably ease the interaction with electronic systems (Benson, 1989). Direct manipulation has become the method by which most computer users interact with their machines. Since the introduction of the Mac GUI in Lisa and the early Macintosh machines, and slightly later the Windows interface, users have come to expect a mouse with visual and physical interaction with their operating system and software. Direct manipulation is evident in many other areas as well. An example of direct-manipulation is resizing a window by dragging its corners or edges with a mouse, or erasing a file by dragging it to the recycle bin.



Direct manipulation has been extensively used in 2D and 3D CAD. A basic approach is presented in (Tonouchi, 1992). Here a framework with whom designers can create visual objects by arbitrarily composing a set of primitives by direct manipulation is described. Virtual GIS is a system for navigating and understanding complex and dynamic terrain-based databases (Koller, 1995). Both window-based and virtual reality versions with direct manipulation capabilities have been implemented. In (Liu, 2008) a method that uses 2D sectional outline curve to drive 3D mesh model deformation is presented. The method uses direct manipulation of free form deformation in order to be more intuitive and real-time. With the development gesture recognition and hand tracking techniques, direct manipulation could benefit of novel techniques to enhance usability of electronic systems. Multitouch interfaces are spreading in modern mobile phones (iPhone). In (Michihiko, 2007) the authors propose a direct manipulation technique for 3D virtual objects. Manipulation is performed by a human playing a role of an actor that wear a glove with whom interact with the virtual objects. Tables provide a large and natural interface for supporting direct manipulation of visual content for human-to-human interactions. Such surfaces also support collaboration, coordination, and parallel problem solving. The SensitiveTable can sense multiple points of contact on surfaces of different shape and size, where gestures can be recognized and become expressive actions (Natural Interaction). An interactive workspace featuring vision based gesture recognition that allows multiple users to collaborate in the creation of a concept map is presented in (Baraldi, 2006). Here the users can browse and modify the contents of a concept maps through a set of simple gestures. Finally, in (Andersen, 2006) a tangible user interface for direct manipulation of sound during playback is presented. The interface allow the mapping of a time varying audio parameter to a tangible handle position. When holding or moving the handle the audio parameter changes and audio playback of the loop is affected instantly.



References

Andersen T.H., et. al (2006). Feel the beat: direct manipulation of sound during playback. First

IEEE International Workshop on Horizontal Interactive Human-Computer Systems.

Baraldi S., et al. (2006). wikiTable: finger-driven interaction for collaborative knowledge-building workspaces. In Proceedings of the 2nd IEEE Workshop on Vision for Human Computer Interaction (V4HCI) in conjunction with IEEE CVPR 2006. New York.

Benson C.R., et al.. (1989). Effectiveness of direct manipulation interaction in the supervisory control of FMS part movement. IEEE International Conference on Systems, Man and Cybernetics. pp.947-952 vol.3.

iPhone, <http://www.apple.com/iphone/features/multitouch.html>

Koller D., et al (1995). Virtual GIS: a real-time 3D geographic information system. Proceedings of IEEE Conference on Visualization. pp.94-100.

Liu B. and Shangguan N. (2008). Constrained Free Form Deformation Driven by Sectional Outline Curve. International Symposium on Computational Intelligence and Design. ISCID '08. vol.2, pp.283-286.

Michihiko, M. (2007). Direct Manipulation of 3D Virtual Objects by Actors for Recording Live Video Content. Second International Conference on Informatics Research for Development of Knowledge Society Infrastructure. ICKS 2007. pp.11-18

Natural Interaction, <http://www.naturalinteraction.org/>

Shneiderman, B. (1983). Direct Manipulation: A Step Beyond Programming Languages. Computer. vol.16, no.8, pp.57-69.

Tonouchi T, et al. (1992). Creating visual objects by direct manipulation. IEEE Workshop on Visual Languages. pp.95-101.

3.3.3.2 PickAndDrop

Pick and drop is an extension of the drag and drop interaction technique that allows the intuitive transfer of objects also across the boundaries of computer systems. Rather than working with a mouse, Pick-and-drop works with a pen such as those made by Wacom and others, as it builds on detecting a hovering (proximity to screen without touching it).

In contrast to Drag and drop, where objects are moved from application to application in a single computer by 'sliding' the object from one position to another while keeping the mouse button pressed, pick and drop works by picking up objects with a pen by single touch and a leaving the screen (finish hovering) on the same place. Then, users can drop the picked-up object at another location by touching at the same spot where the hovering pen was first detected. Thus, for the user it is possible to move objects from one computer's screen to another computer's screen with the same simple gesture just like working on a single screen. For the different computers, the pens are detected by their ID that is normally only considered by a single computer for detecting different tools (e.g. different brushes used for painting).

While conceptually, the user view is that of storing the digital object in the pen (similar to picking up food with a chopstick or fork), the actual transfer is done directly from computer to computer over the network behind the scenes.

References

Jun Rekimoto (1997) "Pick-and-Drop: A Direct Manipulation Technique for Multiple Computer Environments", Proceedings of UIST'97, pp. 31-39, 1997.

3.3.3.3 CrossingUI

Crossing is an alternative to clicking. While in WIMP style user interfaces a mouse click is the fundamental "atomar" operation, crossing a line by the pointing device can be used instead of the click. Accot and Zhai (2002) compared the performance of crossing versus clicking for various tasks. Complete applications such as a painting applications can be based on crossing instead of clicking (ApitzGuimbretiere2004). Crossing is considered especially suited for pen based interaction, e.g. on tablet PCs or interactive whiteboards.

References

Accot, J. and Zhai, S. (2002) More than dotting the i's --- foundations for crossing-based interfaces. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Changing Our World, Changing Ourselves (Minneapolis, Minnesota, USA, April 20 - 25, 2002). CHI '02. ACM, New York, NY, 73-80. DOI= <http://doi.acm.org/10.1145/503376.503390>

Apitz, G., Guimbretiere, F. (2004) CrossY: A crossing based drawing application. UIST 2004, 3-12.

3.3.3.4 Camera Pose Estimation for Augmented Reality

The ability for an Augmented Reality (AR) system to deliver spatially coherent information relies on a key computer vision functionality, referred to as pose estimation. It consists in recovering the position and orientation of the camera with respect to a reference 3D world coordinate system. Several effective algorithms for the pose estimation problem have been proposed in the computer vision literature (e.g. [1], [2], [3]).

Yet, all such algorithms rely on tracking accurately and reliably visual features corresponding to known world objects, which indeed is a very challenging task. As a consequence, many successful AR systems rely on instrumentation of the environment with planar fiducial markers (see [4] for a survey on industrial AR projects), so as to cast the problem into a simpler form that requires tracking highly distinctive planar patterns. To this purpose, specific set of markers and corresponding detection algorithms have been developed, including ARToolKit [5], CyberCode [6] and the circular fiducials proposed by Intersense [7].

However, since instrumentation of the environment with fiducial markers may not be feasible or easily acceptable in a variety of potential AR application scenarios, a significant amount of research has been focused on the development of markerless systems. In such a field, a major issue concerns effective real-time tracking of natural features, i.e. visual patterns that are not stuck on physical objects for tracking purposes but instead do naturally exist in the scene and can be effectively detected and matched into the incoming video-stream in order to provide the required input data to the pose estimation algorithm. In the last years, research on the so-called local features has gained momentum, thanks also to successful projects within the EU FP5 and FP6 programs. Such features consist of visual patterns (e.g. patches, circular blobs, arbitrarily shaped regions, as illustrated in a recent thorough survey [8]) that can be detected and matched in natural images and exhibit

invariance -or robustness- to large scale and viewpoint changes, as well as to image brightness variations. Local features have proven effective to perform computer vision tasks such as object recognition and categorization [9] and have also been deployed for tracking natural patterns in markerless AR systems [3], [10],[11],[12], [13],[14], [15], [16].

References

- [1] Gerald Schweighofer and Axel Pinz. Robust pose estimation from a planar target. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(12):2024–2030, 2006.
- [2] F.Moreno-Noguer, V.Lepetit, and P.Fua. Accurate non-iterative $O(n)$ solution to the pnp problem. In *IEEE Intl. Conf. on Computer Vision*, Rio de Janeiro, Brazil, October 2007.
- [3] Gilles Simon, Andrew W. Fitzgibbon, and Andrew Zisserman. Markerless tracking using planar structures in the scene. In *Proc. of ISAR*, pages 120–128, Munich, Germany, May–June 2000.
- [4] Holger Regenbrecht, Gregory Baratoff, Wilhelm Wilke, *Augmented Reality Projects in the Automotive and Aerospace Industries*, *IEEE Computer Graphics and Applications*, pages 48-56, November/December 2005.
- [5] Hirokazu Kato, Mark Billinghurst, *Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System*, 2nd IEEE and ACM International Workshop on Augmented Reality, 1999.
- [6] Jun Rekimoto, Yuji Ayatsuka, *CyberCode: Designing Augmented Reality Environments with Visual Tags*. *Proceedings of DARE 2000 (Designing augmented reality environments)*, pages 1-10, April 2000, Elsinore, Denmark.
- [7] L Naimark, E Foxlin, *Circular data matrix fiducial system and robust image processing for a wearable vision-inertial self-tracker* , *International Symposium on Mixed and Augmented Reality, ISMAR 2002*, 2002.
- [8] Tinne Tuytelaars, Krystian Mikolajczyk, *Local Invariant Feature Detectors: A Survey*, *Foundations and Trends in Computer Graphics and Vision*, v.3 n.3, p.177-280, 2008. [9] A. Pinz, *Object Categorization*, *Foundations and Trends in Computer Graphics and Vision*, 1(4), pp. 255-353, 2006.
- [10] Iryna Gordon and David G. Lowe, "What and where: 3D object recognition with accurate pose," in *Toward Category-Level Object Recognition*, eds. J. Ponce, M. Hebert, C. Schmid, and A. Zisserman, (Springer-Verlag, 2006), pp. 67-82.
- [11] C. Scherrer, J. Pilet, P. Fua and V. Lepetit, *The Haunted Book*, *International Symposium on Mixed and Augmented Reality*, Cambridge, England, 2008.
- [12] V. Lepetit, P. Laguerre and P. Fua, *Randomized Trees for Real-Time Keypoint Recognition*, *Conference on Computer Vision and Pattern Recognition*, San Diego, CA, June 2005.
- [13] Yuan M.L., Ong S.K. and Nee A.Y.C., "Registration Using Natural features for Augmented Reality Systems", *IEEE Transactions on Visualization and Computer Graphics*, 12(4), 2006, 569-580.
- [14] Gordon I. and Lowe D.G., "Scene Modeling, Recognition and Tracking with Invariant Image

Features”, Proceedings of the IEEE International Symposium on Mixed and Augmented Reality, 2004, DC, USA, 110-119.

[15] Vacchetti L., Lepetit V. and Fua P., “Stable Real-time 3D Tracking Using Online and Offline Information”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(10), 2004, 1385-1391.

[16] P. Azzari, L. Di Stefano, F. Tombari, S. Mattoccia. Markerless augmented reality using image mosaics. In Proc. of ICISP 2008, pages , July 2008.

3.4 Auditory

Auditory-based interaction has been mainly restricted to signaling of feedback or alarms for a long time, as speech based interaction was too difficult or costly to implement. With the diminishing storage constraints allowing for a richer auditory design of user interfaces and increased performance of automatic speech recognition and speech synthesis, auditory-based interaction is gaining popularity.

3.4.1 Input

3.4.1.1 AutomaticSpeechRecognition

In general words, Automatic Speech Recognition (ASR) is the process of algorithmically inferring which words have been spoken by a speaker from recorded speech signals.

State-of-the-art speech recognizers are based on a statistical approach cf. e.g. (RabJua93). Every 10 msec a feature vector encoding spectral information about the speech signal is computed. This builds the basis for the actual word recognition, for which Hidden Markov Models (HMMs) are used. These are state-transition automata, which define doubly embedded stochastic processes. It is assumed that with each state transition one of the above mentioned feature vectors is produced according to a certain probability. Alternative state transitions can be followed according to transition probabilities. Given a feature vector sequence the probability can be computed for a particular HMM having generated this sequence. However, it is unknown (hidden) which state sequence has actually generated this sequence. This approach allows for flexibility in the models which is needed to cope with acoustical and pronunciation variabilities in speech.

Basic HMMs correspond to phonemes. Usually per phoneme several models are used which depend on the phonemic context, so called triphones (or even pentaphones). Word models are built by concatenating these basic HMMs according to a pronunciation lexicon, which contains one or more alternative phoneme sequences per word. A typical speech recognizer contains ten thousands up to millions of statistical parameters.

The parameters/probability distributions of the HMMs have to be estimated using large amounts of training data typically from hundreds of speakers. This is a collection of utterances annotated with the spoken words. For initialization purposes some of the utterances have to be manually segmented into phonemes.

In isolated word recognition (for example, command&control tasks) the task of speech recognition is to find the best matching word level HMM given a feature vector sequence computed out of a

speech signal. In continuous speech recognition the optimal sequence of word HMMs has to be determined. This cannot be done without using a grammar restricting the possible word sequences or sentences. In tasks with limited number of variance in the wording, often hand-coded finite-state-grammars (FSG) are used. In larger tasks stochastic language models are used which contain so-called n-gram probabilities that are probabilities for n-tuples of words. These have to be optimized on large text corpora. Note that all speech recognizers can only recognize words they contain in the lexicon. Generic recognizers able to recognize arbitrary words are not possible with the current state-of-the-art, and it cannot be foreseen when or if at all this will be possible.

In present applications or products speech recognizers are tuned a priori. Tuning here means that the statistical parameters are trained on a task-specific (with regard to vocabulary and environment) speech database, and that lexicon and grammar are developed before the recognizer is released. Afterwards everything is frozen, i.e., while the application is being used, no change in particular of lexicon or grammar is possible anymore. There are research prototypes available that overcome this limitation, however, commercially available recognizers are generally still restricted in this respect.

Speech recognition has benefited from advancements in computing hardware in recent years, as with more memory and computing power it is possible to more thoroughly go through the vast search space in a given timespan and more reliably produce the best-fitting hypothesis. Also restrictions on storing sufficient statistical parameters on devices are more and more alleviated. Finally, at least for the common languages, the manufacturers of speech recognition systems have collected large amounts of speech data - with more and more speech data, the statistical parameters can be estimated more reliably which yields a better recognition accuracy.

Due to the statistical nature and in line with all pattern recognition algorithms, there will always be errors with speech recognition. Obviously, there are more difficult and more easy tasks. In general, recognition becomes harder the more words are to be discerned and the more similar the words are. Thus the task in designing recognition systems is to make sure that users use commands that are phonetically easily discernible. A selection among hundreds of longer phrases can be much easier than the differentiation among ten short digits, for example. For many languages, continuous digit recognition as would be required for general phone dialing is still a very hard task.

Comparably small vocabulary 'command and control' tasks can be considered as solved for quiet environments. Also continuous speech dictation tasks work reasonably well for most speakers in quiet environments. What is a challenge still for speech recognition systems is recognition in the presence of noise, especially speech-like noise, such as from simultaneously speaking subjects. Also truly spontaneous speech with lots of hesitations and clitics, unknown vocabulary as in sociolects and the like pose problems. As the process of training the HMM is relying on large speech databases, for many minority languages or strong dialects, ASR-system are often not available.

References

L. Rabiner and B.-H. Juang. Fundamentals of Speech Recognition. Prentice Hall, New Jersey, 1993.

3.4.1.2 Computational auditory scene analysis

Computational auditory scene analysis technology uses microphones or microphone arrays to

analyse the context by identifying sound patterns. Like computer vision it is applicable in many domains, ranging from public spaces (aggression detection at trainstations for example) to home. It also has similar disadvantages as computer vision: privacy risks, although less severe than for vision, and in principle sensitive to changing acoustic conditions, but that occurs less frequently than changing light conditions. It requires training and extensive databases of representative sound patterns. Computational auditory scene analysis (CASA) is concerned with the research and development of algorithms that attempt to mimic the function of the human auditory system (Wang and Brown, 2007). Often, CASA is used to improve speech recognition in a free acoustic field, to automate score transcription of a music performance, and to enhance intelligibility in a hearing aid. An ambiance provides a mixture of sounds that needs to be sorted out into different streams or events in which each stream or event has acoustical evidence to originate from a single source. With a limited number of microphones and a potential large number of audio-producing sources, this requires a blind separation of the sound mixture into its possible source constituents. The influence of noise, reverb effects, mixes arisen from other concurrent events and various distortions are critical for a successful employment of auditory scene analysis in real practice. Unified approaches to CASA to implement the auditory grouping principles of Bregman have been proposed. The ‘data-driven’ model (Brown, 1992), as an exemplar of the many ‘bottom-up’ approaches, computes, for each small time slice, a set of acoustical cues from the complex sound mixture such as periodic modulation, spectral content (by means of MFCC), transients, pitch structure, onsets and offsets. These cues are placed in a representational framework to form acoustic objects. These objects are then used in grouping algorithms to collect those objects that come from the same source. However, the basic criticism of these ‘bottom-up’ systems is simply that they do not work. The expected scope of use of these models was over-ambitious. They are not equipped to model application specific details. For instance, they do not model the expectation of sound occurrence in an environment. An alternative is the ‘ecological-prediction-based’ approach, inspired by the work of Gaver (1993) and Ellis (1996), in which the observed acoustical cues are constantly reconciled with the predictions of a model of the sound-producing entities in the environment. In this way, the acoustic scene is partly interpreted by means of predicted sound events, to measure whether or not there is direct acoustic evidence for them. Also, the ambiance can be explicitly modeled, in a probabilistic sense, which anticipates what one is likely to expect. This approach has been applied for automatic speech recognition with interfering sources by e.g. Barker et al., 2005.

References

- Bregman, A.S. (1990). *Auditory Scene Analysis: The perceptual organization of sound*. Cambridge, Massachusetts: The MIT Press, 1990.
- Van Noorden, L.P.A.S. (1975). *Temporal Coherence in the Perception of Tone Sequences*,” Ph.D., Technische Hogeschool Eindhoven, The Netherlands.
- Wang, D., and Brown, G.J. (Eds.) (2006). *Computational Auditory Scene Analysis: Principles, Algorithms and Applications*. Wiley-IEEE Press.
- Brown, G.J. (1992). *Computational auditory scene analysis: A representational approach*, PhD. Thesis, CS-92-22, CS dept, University of Sheffield.
- Ellis, D.P.W. (1996). *Prediction-driven computational auditory scene analysis*, PhD. Thesis, Dept.

of Electrical Engineering and Computer Science. Massachusetts Institute of Technology.

Gaver, W. (1993). What in the world do we hear?: An ecological approach to auditory event perception, *Ecological Psychology*, 5, 1, 1-29.

Barker, J.P., Cooke, M.P., and Ellis, D.P.W. (2005), Decoding speech in the presence of other sources. *Speech Communication*, 45, 5-25

3.4.2 Output

3.4.2.1 SpeechSynthesis

Speech synthesis is the process of generating audible speech signals from symbolic representations of speech, such as texts. A simple playback of previously recorded voice messages is generally not considered as speech synthesis, although it could be seen as a degenerated extreme case of concatenative speech synthesis.

The process of speech synthesis of an utterance generally involves an analysis of the text, a conversion to phonemes to generate the pronunciation, and the actual generation of speech signals.

There are different techniques for generating the speech signals: formant synthesis, diphone synthesis and unit selection synthesis, with diphone synthesis being used most frequently used, and unit selection synthesis promising most natural sounding synthesis.

References

Alan W. Black and Paul A. Taylor. CHATR: A generic speech synthesis system. In COLING '94, Kyoto, Japan, 1994.

3.4.2.2 Nonspeech Audio

Sound plays an integral role in people's everyday life and in how they perceive the world around them. People are surrounded by sounds, some are helpful, while others may be not. The fact is that in daily life people make use of all their senses simultaneously and therefore interacting with a system in a multi-modal way is considered to be more natural. In human communication speech interaction is the most common way, but people are also very good in assessing other types of sound. For instance, when they cross a street, are cooking or hear a walking person approaching, people use nonspeech audio cues to determine what is happening around them. This means that the skills that users have built up over a lifetime of everyday listening can be exploited in user-system interaction (Gaver, 1986) (Buxton, 1989).

One benefit of sound is that audio messages are received regardless of where the user is looking, it does not require continuous visual monitoring. A very important function of sound is also to provide redundant information to users (Gaver, 1986) as this increases the feeling of direct engagement items represented in a user interface.

Different types of nonspeech audio can be distinguished. Some nonspeech audio cues are building on existing everyday knowledge (called "auditory icons") whereas others employ a newly designed vocabulary (called "earcons"). If chosen correctly, the use of auditory icons requires less learning from the part of the user because they refer to everyday sounds. Earcons use a more musical

approach, and are often constructed of synthetic tones.

References

Buxton, W. (1989), Introduction to this special issue on nonspeech audio, *Human-Computer Interaction*, 4, 1-9.

Gaver, W.W. (1986), Auditory icons: using sound in computer interfaces, *Human-Computer Interaction*, 2, 167-177.

Van de Sluis, Eggen, Rypkema (1998), Nonspeech Audio in Television User Interfaces, in proceedings of HCI'98, Sheffield.

3.5 Haptic (touch and proprioception)

Whilst most output based interaction technologies are vision-based in nature, for the input side, haptics is by far the predominant modality. Typing with a keyboard or keypad, pointing with a mouse, cursor, joystick, pen or finger are probably the most widespread ways of input today. All of these are building on our ability to operate very well with our hands and fingers.

3.5.1 Input

3.5.1.1 Typing

A computer keyboard is modeled after the typewriter keyboard, which uses an arrangement of keys, and is mostly used for text input.

3.5.1.2 Alternative Text Input

There is a multitude of alternative text input methods that have specifically been designed for mobile devices or generally for situations where a standard keyboard is too bulky. Many of them work with the numeric keypad found on mobile phones. We pick some of the more well known or influential technologies and two techniques available inside the consortium.

Several of the methods utilize redundancies in natural language, i.e. by utilizing statistical properties of the language, input is simplified for the user (in terms of reduced button presses or reduced travel of a pen or so). These methods are frequently called 'predictive'.

3.5.1.2.1 Methods utilizing numeric keypad

The International Telecommunication Union specifies in standard E.161 the arrangement of letters on digit keys for phones.

Multitap is the standard text input method found on almost every phone where the nth letter on any button is produced by n-times pushing this button in sequence. Multitap is criticized for the usage of timeouts that are required to allow input of successive letters on the same key. These timeouts trouble first-time users that must learn the timespan, as well as frequent users, that are limited in typing speed by the pauses (see, e.g. McKenzie04).

Tegic's T9 is the most widespread predictive text system today. Ideally, for each letter, only one

keypress is required, language statistics chooses the right word. In practice, T9 is more difficult to operate, as seemingly correct word prefixes can change later, the system is problematic in the case of proper nouns, and if the users can not reach the word they wanted to enter by T9's prediction, they have to completely go back and use multitap (and end up with many more keypresses and with increased frustration). Still, for many users, T9 is helping in inputting text more quickly.

Eatoni's Letterwise is a system comparable to T9 that tries to get around some of T9's and Multitap's shortcomings. The inventors have shown the superiority over T9 and Multitap in (McKenzie&al2001), and according to Eatoni's web site there are several DECT and mobile phones with a preinstalled Letterwise.

TiltText disambiguates which of the letters are chosen for a button by asking the user to tilt the device simultaneously into a specific direction (left for A, top for B right for C while pressing 2 etc.) An advantage of TiltText is that no timeouts are required and thus operation without visual attention is possible after the user has remembered the key's mapping (WigdorBalakrishnan2003).

3.5.1.2.2 Variants of qwerty

As many people are familiar with the layout of a qwerty keyboard, some approaches try to leverage this knowledge. Visual on-screen keyboards can be drawn on a screen and operated by a pen or even fingers. As a disadvantage, visual keyboards consume rather large amounts of the scarce screen real estate and if made smaller, are error-prone in operation. Also typing is more difficult as it lacks a haptic feedback (but see Poupyrev&al2002).

A solution to have a large keyboard off-screen without the required space for moving the device is the virtual keyboard by Canesta and others. They project a keyboard on any suitable flat surface and recognizes input by a (3D-time-of-flight) camera. The same disadvantage of missing haptic feedback holds also in this case.

3.5.1.2.3 Chorded text entry

There are various methods for typed input using simultaneous keypresses, in general requiring 5 to 8 keys only (typically one for each finger except for the thumb that chooses among several). The main advantage is that it allows for quick one-handed text input. Application has been restricted to professional use and as an aid for handicapped persons, as training is required to memorize the chords. Chorded text entry has found some attention among proponents of wearable computing (LyonsStarnesGane2006).

There is also a technique that lies inbetween the qwerty and the chorded approach: the half-qwerty (Matias&al1996). Here, the space bar is used to toggle between left and right side of a standard keyboard so that one-handed input on a half keyboard is possible.

3.5.1.2.4 Penbased

Some methods have been designed for text entry by pen that is beneficial for tablet PCs and electronic whiteboards.

Quickwriting (Perlin1998) and Cirrin (MankoffAbowd1998) use variants of a PieMenu. Zhai and Kristensson (2003) utilize a unistroke gesture over the keys of an on-screen keyboard. Frequently used words have characteristic traces that can be memorized by skilled users. EdgeWrite

(Wobbrock&al2003) is similar to the unistroke alphabet used in Palm's Grafitti, but takes advantage of edges around the writing area. Thus also people with motor control problems can utilize handwritten input or writing letters can be better accomplished without looking.

3.5.1.2.5 Technologies available in the consortium

Conante has two technologies for text input that are efficient, easy learnable and can be integrated on small displays without wasting screen real estate. Also, they use few input controls.

CharacterPump is a technology that allows text input by the help of a rotary dial that has haptic feedback. Letters are visualized on a vertical ribbon that follows the rotations of the dial. It is a predictive systems in that it uses language statistics to ease access to frequent words by making more frequent letters larger and less frequent smaller so that the frequent letters are more easily reached. The size of the letters can be felt through the haptic feedback (the frequent ones 'snap in' easily). Inputting letters is frequently only a sequence of pushes of the dial, as the most likely letter is preselected after each entry. In contrast to T9, no words are excluded a priori and there is just one mode of operation. No specific assignment of letters to keys or locations of letters has to be learned (video).

EnChoi is a technology comparable to CharacterPump, but it works on standard handset hardware, i.e. also without a haptic dial. The ribbon is moved by strokes on the touchscreen or touchpad, or operated by cursor keys / joystick, as is the entering of the selected letter. EnChoi can also be used to select entries from a large database, the letters that are input work as an incremental filter on the elements of the database, which makes e.g. selecting a song from a music database both easy and efficient.(see also interaction paradigm 'one handed interaction')

References

MacKenzie, I.S., Kober, H., Smith, D., Jones, T., Skepner, E. (2001). LetterWise: Prefix-based disambiguation for mobile text input. ACM UIST Symposium. p. 111-120.

Wigdor, D., Balakrishnan, R. (2003). TiltText: Using tilt for text input to mobile phones. ACM UIST Symposium. p. 81-90.

Poupyrev, I., S. Maruyama, and J. Rekimoto. TouchEngine: A tactile display for handheld devices. Proceedings of CHI 2002, Extended Abstracts. 2002: ACM: pp. 644-645

Lyons, K., Starner, T., and Gane, B. 2006. Experimental evaluations of the Twiddler one-handed chording mobile keyboard. Hum.-Comput. Interact. 21, 4 (Nov. 2006), 343-392.

Edgar Matias, I. Scott MacKenzie, William Buxton (1996) Half-QWERTY: Typing with one hand using your two-handed skills CHI'96 Conference Companion, pp. 51-52, April 1996

Ken Perlin (1998) Quikwriting: Continuous Stylus-based Text Entry, Proceedings of the ACM Symposium on User Interface Software and Technology (UIST'98), pp. 215-216, November 1998

Jennifer Mankoff, Gregory D. Abowd (1998) Cirrin: A word-level unistroke keyboard for pen input Proceedings of the ACM Symposium on User Interface Software and Technology (UIST'98), pp. 213-214, November 1998

Shumin Zhai, Per-Ola Kristensson (2003) Shorthand Writing on Stylus Keyboard Proceedings of

the ACM Conference on Human Factors in Computing Systems (CHI2003), pp. 97-104, April 2003
Jacob O. Wobbrock, Brad A. Myers, John A. Kembel (2003) EdgeWrite: A Stylus-Based Text Entry Method Designed for High Accuracy and Stability of Motion Proceedings of the {ACM} Symposium on User Interface Software and Technology (UIST2003), November 2003

3.5.1.3 Pointing by Mouse or Pen or Finger

In many different shapes, these devices are used in all kinds of applications to sense from 1 up to 6 degrees of freedom (6 = 3 displacement + 3 rotational). Many use mechanical transduction - displacement or acceleration sensing, though optical techniques are also very popular.

3.5.1.4 Writing, Handwriting recognition, gesture recognition

Except from computer keyboards, alternative means exist for textual input. For example digital pens; pens that use a very small trackball in the tip, or inertia sensors, to read hand writing or do signature recognition. Furthermore, pens sold with certain pen tablets use a pressure sensor in the tip to produce more natural writing.

3.5.1.5 SwitchesDialsLevers

Electromechanical controls are utilized in almost all electronic devices, from the very beginnings to current times. Although the importance of electromechanical controls has been reduced by the omnipresence of touchscreens and pointing devices, they still are an important contribution to human computer interaction.

There are various types available, suitable for different applications. Most frequently used switch types include pushbutton, toggle, rocker style, rotary style, slide, DIL, joystick and key operated switches (source www.farnell.com). Rotary dials with optical encoders are also frequently used (mouse, data entry dials). Potentiometers have been widely used to adjust volume or other parameters in a direct way.

Electromechanical controls can also transport subtle information or quality perception to customers (Michelitsch&al04, see HapticUI). It has been noted that purely graphical user interfaces could benefit from physical counterparts of the graphical widgets, so-called phidgets (GreenbergFitchett2001).

References

Greenberg S, Fitchett C (2001) Phidgets: easy development of physical interfaces through physical widgets. In: Proceedings of the 14th annual ACM symposium on user interface software and technology (UIST 2001), Orlando, Florida, November 2001, pp 209–218

3.5.1.6 Directional input / Joysticks, Cursor keys, Jogdials

Especially in the games domain a variety of directional input controllers is available: such as joysticks, steering wheels, pedals, pistols and game pads. Those often use mechanical sensors to sense the movement. Some of these controllers also provide force feedback.

3.5.1.7 Gloves

Gesture interfaces contribute to the feeling of immersion and effectiveness of interaction. One of our primary physical connections to the world is through our hands. We perform most everyday tasks with them. In an effort to increase the naturalness of Human Computer Interaction (HCI), several techniques have been studied to recognize user hand movements. The first steps in hand gesture recognition have been done in the late 70s at MIT with the Put-that-there project (Bolt, 1980). In this project a set of magnetic field sensors were used to retrieve the position and orientation in the space of the user hands. This information was used, together with speech recognition, to select, move and query graphical elements on a large screen. Gloves are cost-affordable devices mainly based on use of off-the-shelf components. Typical sensors used are inertial and bend sensors. However, in general, the majority of commercial devices only handle bending of finger and palm or the finger abduction. CyberGlove (CyberGlove) is an interesting device, equipped with a rich set of bend sensors, monitoring different kinds of finger movements. Three dimensional positioning is obtained through an optional motion tracking module that can be connected to the glove. The second version of the CyberGlove (CyberGloveII) allows the connection of more motion tracking modules and integrates a LED to provide the user with a feedback from the system. A simpler solution is the 5DT Data Glove (5DT Data Glove) and its variants. The basic configuration offers only a subset of functionalities, while the more equipped one can manage abduction between fingers but still does not handle translations. The devices presented above are wired solutions, however an external Bluetooth module can be connected to provide wireless connectivity. A slightly different approach is proposed with the P5 Glove (P5 Glove). The P5 has been specifically designed to enhance the PC game-playing experience. It exploits 5 bend sensors placed on user fingers to track fingers bending while hand tracking is performed through several LEDs placed on the glove and a receiver that tracks their position in the space. The receiver range is limited to 3-4 foot range. Gloves and usage of sensors (accelerometers, gyroscopes and bend sensors) for building interfaces and designing interaction devices is not only interesting for commercial use but it has been largely explored in many research studies since a couple of decades (Zimmerman, 1987). In (Perng, 1999) a system interface dedicated to hand movement recognition to enable mouse-like input using an accelerometer glove has been presented. The glove is equipped with six 2-axis accelerometers on the finger tips and back of the hand. Accelerometers are used as inclinometers to detect 28 static gestures. The glove has also an RF transmitter to send data to a personal computer, thus acting as a wireless input device. The AcceleGlove is a portable low-cost glove based on accelerometers used to recognize the 26 basic gestures of the American Sign Language through a three-level hierarchical classifiers. Similar examples can be found also in (Liang, 1998) and (Waldron, 1995). Another example has been presented in (Farella, 2008). Here a proximity-aware smart wireless glove combines accelerometers and bend sensors and is used as a gesture interface. The system has been used in several applications: 3D game environment, interaction with 3D graphical shapes, music in a MIDI application, as a pointing device in a cultural heritage application (see figure below).



Gloves can be used also as text input devices when using in conjunction with pressure sensors. The chording glove uses chording to generate characters, similar to a chord keyboard, except that the keys are mounted on the fingertips of a glove and the chords are formed by pressing against any hard surface (Rosenberg, 1999). Another solution is proposed in (Shin, 2005). Here one hand of the user acts as the keyboard, and the phalanges of the fingers of it are used as the keys of a phone keypad. As a consequence one person types on his/her hands. A different approach to the bending of fingers is presented in (Fu, 2007), where fingers inclination is detected using LEDs and photo detectors. This device is used as input text device for impaired people. Finally Gloves interfaces can integrate an haptic feedback to the user through a set of motors placed on an exoskeleton around the glove. The CyberGrasp device is a lightweight, force-reflecting exoskeleton that adds resistive force feedback to each finger. With the CyberGrasp force feedback system, users are able to feel the size and shape of computer-generated 3D objects in a simulated virtual world (CyberGrasp). A survey technologies for enhancing tactual feedback source is presented in (Bonger, 1997). Such techniques can be embedded into glove design to provide a more natural interaction. For example the DeKiFeD3 is a 3 degrees-of-freedom haptic interface with the capability to generate kinesthetic hand feedback that has been combined with a commercial haptic glove to generate additional kinesthetic feedback to the fingers of a human operator (Kron, 2000). This device is used in conjunction with visual and audio feedback to provide immersive user experience.

References

5DT Data Glove, http://www.vrlogic.com/html/5dt/5dt_dataglove_5.html

Bolt, R. A., (1980). "put-that-there": Voice and gesture at the graphics interface. In Proceedings of the 7th annual conference on Computer graphics and interactive techniques (SIGGRAPH '80). Vol. 14. ACM Press, pp. 262-270.

Bonger, B. (1997). Tactile display in electronic musical instruments. IEE Colloquium on Developments in Tactile Displays, pp.7/1-7/3.

CyberGlove. <http://www.vrlogic.com/html/immersion/cyberglove.html>

CyberGloveII, http://www.vrlogic.com/html/immersion/cyberglove_ii.html

CyberGrasp, <http://www.vrlogic.com/html/immersion/cybergrasp.html>

Hernandez-Rebollar J.L., et al (2002). A multi-class pattern recognition system for practical finger spelling translation. In 4th International Conference on Multimodal Interfaces (ICMI), pages 185–190.

Kron, A. et al. (2000). Exploration and manipulation of virtual environments using a combined hand and finger force feedback system. Proceedings. 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2000. (IROS 2000). vol.2, no., pp.1328-1333.

Liang R. and Ouhyoung M. (1998). A Real-time Continuous Gesture Recognition System for Sign Language. Proceeding of the Third IEEE Int. Conf. On Automatic Face and Gesture Recognition. pp 558-567.

P5 Glove, <http://www.essentialreality.com/>

Perng J. K., et al. (1999). Acceleration sensing glove (ASG). In 3rd International Symposium on Wearable Computing (ISWC), pages 178–180.

Rosenberg, R. and Slater, M.(1999). The chording glove: a glove-based text input device. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol.29, no.2, pp.186-191.

Shin, J.-H. and Hong, K.-S. (2005). Keypad gloves: glove-based text input device and input method for wearable computers. Electronics Letters, vol.41, no.16, pp. 15-16.

Waldron M.B., et al. (1995). Isolated ASL Sign Recognition System for Deaf Persons. IEEE Trans. On Rehabilitation Engineering, vol 3 No. 3.

Fu Y.-F. and Ho C.-S. (2007). Development of a Programmable Digital Glove. International Conference on Machine Learning and Cybernetics, vol.4, no., pp.1948-1953.

Zimmerman T. G., et al.(1987). A hand gesture interface device. In CHI '87: Proceedings of the SIGCHI/GI conference on Human factors in computing systems and graphics interface, pages 189–192, New York, NY, USA. ACM Press.

3.5.2 Output

3.5.2.1 Tactile feedback

The human haptic sense is composed by two sub-modalities: the kinaesthetic sense (force, motion) and the tactile sense (tact, touch). Focusing only on the second modalities, several works explore touch stimulus on the skin with different frequency and different location in HCI, BCI and healthcare domain. To provide direct stimulus four major kind of mechanoreceptors are involved in function of the sensation to provide [1]:

- Pacinian corpuscle [40-500Hz]: detect acceleration or vibration;
- Ruffini ending [100-500Hz]: provide buzz-like sensation;

- Merkel's disk [0.4-3Hz]: detect force or displacement, spatial information on static condition;
- Meissner corpuscle [2-40Hz]: detect dynamic deformation, flutter sensations;

The application domain are various: teleoperation and telepresence; sensory substitution, integration or augmentation; 3D surface generation; Braille systems; games, etc. Many are the technologies to provide feedback. We can conceptually divide them in two categories: (i) Vibrotactile Stimulators, thought to exploit the persistency of the sensation, stimulating mainly Pacinian corpuscles through playing with the wave length of the vibration more than with its amplitude; they mainly provide information related with the sensation of texture or superficial roughness of a material in contact with the skin; (ii) shape displays that aim at imitating an object shape by use of a certain number of tactors, which stimulate orthogonally (Orthogonal Indentation) or laterally (Lateral Skin Stretch) the most superficial skin layer. Some technologies typically in use are:

- Electromechanical actuators, pneumatic (e.g. inflating balloons [2]) or hydraulic actuators
- Piezoelectric actuators
- Shape Memory Alloys (SMA) [5]
- Dielectric Elastomer Actuators (DEA) or Electro Active Polymers [4]
- MEMS actuators for Tactile Displays
- Vibrating motors (eccentric rotating mass or Linear-Rotary Actuator)
- Solenoids or Voice Coil [3]
- Electrorheological fluids (for providing sensation of variation of viscosity)
- Thermoelectric elements

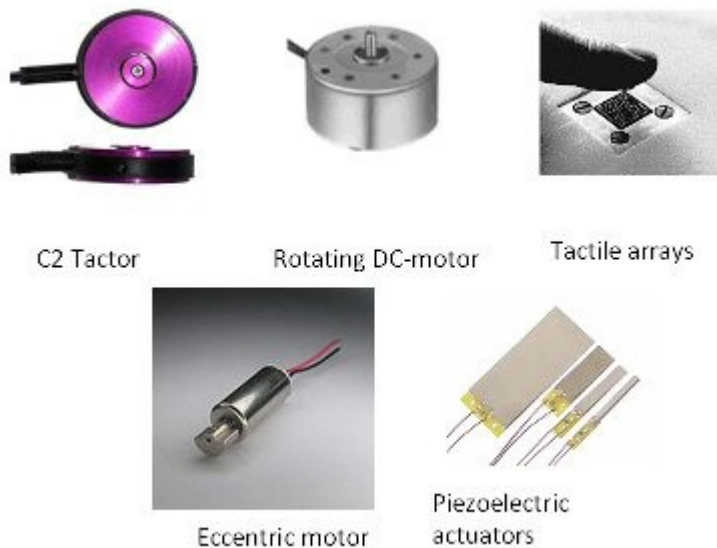


Figure: Example of technologies for tactile feedback

Vibrotactile actuation is commonly used in mobile phones, which is the most widely available type of handheld computer or smart object that each of us uses everyday. The vibrations used in these devices are very simple and do not fully exploit the potential of vibration as a means of communication. However, their potential is clear: they can provide attention grabbing notifications to users engaged in unrelated tasks (which screen-based visual cues cannot) and do this discreetly (which speaker-based audio cues cannot) and without explicitly interrupting users. As mentioned, vibration alerts generally consist of a simple buzz to alert users to incoming calls or messages, or to upcoming calendar appointments. Nevertheless, research on vibrotactile displays for mobile devices has developed and evaluated complex multi-dimensional tactile stimuli with promising results [9]. Tactons [6] are structured, abstract, tactile messages which can be used to communicate information nonvisually. Using Tactons in smart objects, handheld devices and for mobile phone alerts could enable tactile-only communication of complex information. Use of more complex vibrotactile feedback can for example suggest the source and category of the alert and its priority. Moreover, it can significantly improve (i) immersive experiences in virtual environments [7], (ii) teleoperation and telepresence [8] applications, (iii) Sensory substitution (e.g. Braille systems). In combination with audio and visual feedback, vibrotactile stimuli can significantly improve the interaction with the user.

References

- [1] R.S. Johansson, A.B. Vallbo, "Tactile sensibility in human hand: relative and absolute densities of four types of mechanoreceptive units in glabrous skin", 1979, Journal of Physiology, Vol. 286, Page (s): 283-300
- [2] E. Igarashi, K. Sato, M. Kimura, "Development of a Tactile Feedback Device Used a Pneumatic Balloon Actuator", Proc. Of the Second International Symposium on measurement and Control in

Robotics, ISMCR, Tsukuba Science City, Japan, 15-19 November, 1992.

[3] Voice coil actuators for human-robot interaction, McBean, J.; Breazeal, C.; Intelligent Robots and Systems, 2004. (IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on Volume 1, 28 Sept.-2 Oct. 2004 Page(s):852 – 858 vol.1

[4] Micromachined polymer actuators as tactors for tactile display, To, T.K.H.; Zhou, J.W.L.; Hoyin Chan; Li, W.J.; Yunhui Liu, Robotics, Intelligent Systems and Signal Processing, 2003. Proceedings. 2003 IEEE International Conference on Volume 2, Issue , 8-13 Oct. 2003 Page(s): 704 - 709 vol.2

[5] Tactile display for 2-D and 3-D shape expression using SMA micro actuators, Matsunaga, T.; Totsu, K.; Esashi, M.; Haga, Y. Microtechnology in Medicine and Biology, 2005. 3rd IEEE/EMBS Special Topic Conference on Volume , Issue , 12-15 May 2005 Page(s): 88 – 91

[6] Brown, L.M., Brewster, S.A., and Purchase, H.C. A First Investigation into the Effectiveness of Tactons, in Proc. World Haptics 2005, IEEE Press (2005), 167-176.

[7] Benali Khoudja M., Hafez M., Alexandre JM. et Kheddar A. “Tactile Interfaces: A state of the Art Survey” International Symposium on Robotics (ISR 2004), Paris, France, 24-26 Mars 2004.

[8] D.G. Caldwell, C. Gosney, "Multi-Modal Tactile Sensing and Feedback (Tele-Taction) for Enhanced Tele-Manipulator Control", Proceedings of the IEEE/RSJ, Yokohama, Japan, July 26-30, (1993).

[9] Williamson, J., Murray-Smith, R., Hughes, S., 2007. Shoogle: multimodal excitatory interaction on mobile devices. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'07). ACM Press, pp. 121–124.

3.5.3 INPUT/OUTPUT

3.5.3.1 Smart Objects

Ubiquitous computing involves integrating tiny microelectronic processors, communication units and sensors into everyday objects in order to make them smart. Smart things can detect their environment and their appearance tie them to a specific duty, therefore they helps users to cope with their tasks in new, intuitive ways. The domain of Smart Objects is vast. As sensors for light, pressure, temperature, vibration, actuators, and other similar objects evolve, new applications and solutions are being created and implemented.

Several research groups are developing smart object. Since 1995 the Things That Think (TTT) consortium at MIT Media Lab began with the goal of embedding computation into everyday things such as clothing, jewelry, and tables (Things that Think, 1995). For example the I/O brush is a drawing tool able to detect colours and texture of physical surfaces using a CCD camera. These are used as the new "ink" to draw on a large touch screen with a back projection screen (Ryokai, 2005). The Research Group Embedded Interaction, founded in 2004, develops concepts, methods and toolkits to advance embedded human computer interaction. This group developed a generic wireless sensor node called Smart-Its that exploit several kind of sensors and can be embedded into different

object to augment them with embedded processing and interaction capabilities (Gellersen, 2004). This platform has been used in several work to explore how augmented artifacts can cooperatively reason about their situation in the world (Strohbach, 2004), to establish digital links among different personal artifacts (Holmquist, 2001) and to track activity through weighting surfaces (Schmidt, 2003). A nice example of smart object is the MediaCup, an ordinary coffee cup invisibly augmented with sensors, processing and communication (Beigl, 2001) (see figure below). The cup is able to sense its movement and temperature and to share these information to support context-aware systems.



Figure: MIT media Cup

Wireless sensor nodes can be used to augment traditional electronic devices such as smart phones in order to react to their context in a smart manner. For example the project Technology Enabling Awareness (TEA) developed a multisensor module to provide context-awareness to mobile phones. Context-awareness is used, for example, to identify among the situations in-hand, on-table, in-pocket, and outdoors and adapt the profile of the phone to its situation (Schmidt, 2001). Smart object are used to improve social interaction. For example, the Lover's Cups developed at MIT Media Lab are sensor augmented cups able recognize if they are in use (Chung, 2006). Cups are linked in couple, when both cups are used at the same time they provide a visual feedback to the user in order to reinforce social ties between people. Another example are the ComSlippers developed at Carnegie Mellon University. ComSlippers are sensor augmented slippers with whom an user can communicate his/her emotion to his/her partner by performing some specific movements (Chen, 2006). Smart object can be built by embedding computational and communication capabilities into furniture and household appliances. The Nutrition-Aware Kitchen developed at National Taiwan University recognizes cooking activities and the amount of food used

through weighting sensors in order to provide nutritional information (Chi, 2007). A similar approach has been used for the Dining Table developed also at National Taiwan University (Chang, 2006). The Internet Fridge from LG embeds a computer with a LCD display mounted on the fridge door for Internet and multimedia access as well as for other appliances control (LG, 2002). Smart object carried or worn by the user can be used to perform activity recognition in an unobtrusive manner by analyzing how they are used. For example, the Intelligent Gadget is a smart object for personal life logging (Kugsang, 2008). The Intelligent Gadget has limited resources in processing and power, thus it implements a simple activity recognition techniques with an hierarchical structure. In (Joguet, 2003) a pen like smart object able to recognize handwriting has been developed. The pen do not need any tablet or special paper to operate and can be used for transcription or for signature verification. The Sensor Button is a wearable sensor node with the form factor of a button that can be unobtrusively integrated into user garment to perform activity recognition (Roggen, 2006). The Sensor Button integrates an accelerometer, a microphone and a light sensor together with a 16 bits microcontrollers and a wireless transceiver. Several classification algorithm have been implemented on it to recognize gestures in a workshop scenario, in the use of household appliances scenario or in a office scenario. Smart objects are a natural choice to augment the power of Tangible User Interfaces (TUIs) since on board processing capabilities allow autonomous recognition of the activity that the user performs with them (Thompson, 2005). Furthermore, being wireless they can be carried around thus increasing the flexibility of the system. TANGerINE is a tangible tabletop environment where users manipulate smart objects in order to perform actions on the contents of a digital media table. Unlike other approaches, here the user manipulate the SMCube (Cafini, 2008) a smart object that allows interaction both on the tabletop and in the area nearby the table. Smart objects have proven to be a valid tool for entertainment and education. For example, in (Kranz, 2006) a cube shaped smart object with displays on its faces has been used by children to perform a set of multiple choices tests. Children were able to quickly understand how the object works and the experience improved the collaborative aspect of learning. In (Baraldi, 2008) two application scenarios are presented for the SMCube (see figure below). In the TANGerINE Theatre the audience of a theater play can interact with actor by changing the background of the performance, actors improvise adapting the stories to the different changing settings. TANGerINE Tales is an application used to support multiple users face-to-face collaborative story macking. Another smart device used to study children collaborative story making is TellTale, a caterpillar-like toy with five modular body pieces and a head (Ananny, 2001). On each body piece there is a button that allow the recording of 20 second of speech. Body parts can be rearranged in different sequences to build several stories.

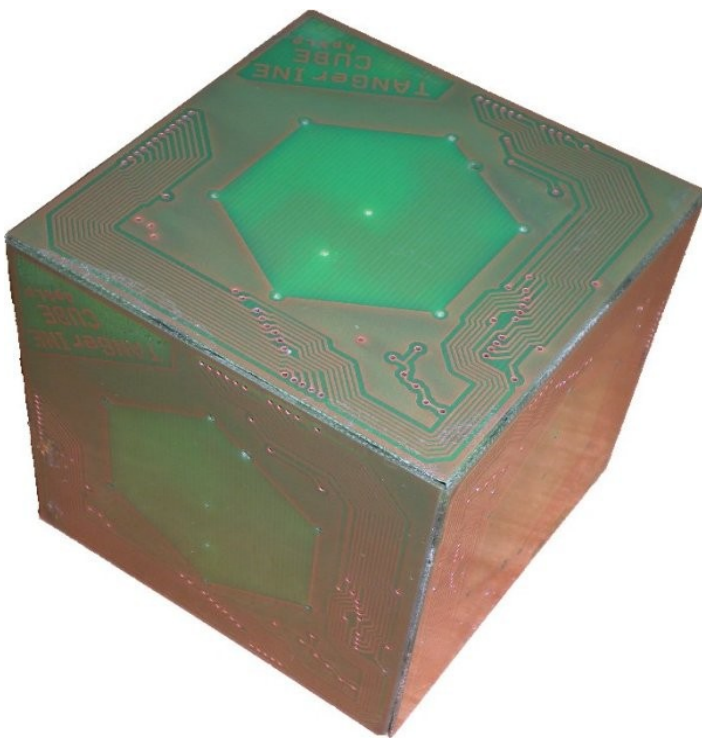


Figure: The Smart Micrel Cube

References

- Ananny M., and Cassell J. (2001). TellTale: A toy to encourage written literacy skills through oral storytelling. In Winter Conference on Text, Discourse & Cognition, pp.28-30.
- Baraldi S., et al. (2008). Evolving TUIs with Smart Objects for multi-context interaction. In Work In Progress proceedings of CHI 2008, Florence.
- Beigl M., et al. (2001). Mediacups: experience with design and use of computer-augmented everyday artifacts. Computer Networks (Amsterdam, Netherlands: 1999) Vol. 35, n.4, pp. 401-409.
- Cafini O., et al. (2008). Tangerine SMCube: a smart device for human computer interaction. Adjunct Proceeding of the 3rd European Conference on Smart Sensing and Context (EuroSSC), Zurich, CH, p.23-24.
- Chang K.-H., et al.(2006). Dietary-Aware Dining Table: Observing dietary behaviors over tabletop surface. In Proceedings of the 4th International Conference on Pervasive Computing (Pervasive 2006).
- Chen C. -Y., Forlizzi J., Jennings P.(2006). ComSlipper: An Expressive Design to Support Awareness and Availability. Alt.CHI Paper in Extended Abstracts of Computer Human Interaction (ACM CHI 2006).
- Chi P.-Y, Chen J.-H., Chu H.-H.(2007). Enabling Nutrition-Aware Cooking in a Smart Kitchen. Work-in-Progress Paper in Extended Abstracts of Computer Human Interaction (ACM CHI 2007).
- Chung H., Lee C.H., Selker T.(2006). Lover's Cups: Drinking Interfaces as New Communication

- Channels. CHI Paper in the Extended Abstracts of Computer Human Interaction (ACM CHI 2006).
- Gellersen H., et al. (2004). Physical prototyping with Smart-Its. *IEEE Pervasive Computing*, vol.3, no.3, pp. 74-82.
- Holmquist L.E. , et al.(2001). Smart-Its Friends: A Technique for Users to Easily Establish Connections between Smart Artifacts. *Proc. 3rd Int'l Conf. Ubiquitous Computing (UbiComp 01)*, pp. 116–122.
- Joguet C., Caritu Y. and D. David (2003). Pen-like, Natural Graphic Gesture Capture Disposal, Based on a Microsystem. In *Proc. Smart Object Conference*, Grenoble, France.
- Kranz M., et al. (2006). The Display Cube as Playful TUI To Support Learning. Video submission at Pervasive 2006.
- Kugsang J. et al. (2008). User activity recognition and logging in distributed Intelligent Gadgets. *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems 2008 (MFI 2008)*, pp.683-686.
- LG (2002). GRD-267DTU Digital Multimedia Side-By-Side Fridge Freezer with LCD Display. <http://www.lginternetfamily.co.uk/fridge.asp>.
- Ryokai K., Marti S., Ishii, H. (2005). Designing the World as Your Palette. In *Proceedings of Conference on Human Factors in Computing Systems (CHI '05)*, Portland, OR.
- Roggen D., et al. (2006). From Sensors to Miniature Networked SensorButtons. In *Proc. 3rd International Conference on Networked Sensing Systems (INSS)*, Chicago, USA.
- Schmidt A. and Van Laerhoven K. (2001) How to Build Smart Appliances? *IEEE Personal Communications*. Vol. 8, n.4.
- Schmidt A., et al.(2003). Ubiquitous Interaction: Using Surfaces in Everyday Environments as Pointing Devices. *Proc. 7th ERCIM Workshop User Interfaces for All (UI4ALL 02)*, pp. 263–279.
- Strohbach M, et al.(2004). Cooperative Artefacts: Assessing Real World Situations with Embedded Technology. *Proc. 6th Int'l Conf. Ubiquitous Computing (UbiComp 2004)*, pp. 249–266.
- Thing That Think (1995). MIT. <http://ttd.media.mit.edu/index.html>
- Thompson, C.W. (2005). Smart devices and soft controllers. *IEEE Internet Computing*, vol.9, no.1, pp. 82-85.

3.5.3.2 Force feedback

"Force feedback interfaces can be viewed as computer “extensions” that apply physical forces, and torques on the user." (Burdea1999). Force feedback has a tradition in teleoperation already in the 1950ies and 1960ies, used for handling of nuclear material from a distance to avoid a nuclear exposure to the operators. Force feedback had a renaissance in the early 1990ies, where gloves with haptic feedback and desktop haptic devices such as the SensAble Phantom have been developed.

When a haptic feedback force is generated, it has to be done more quickly than with visual systems to avoid unwanted oscillations and to be haptically plausible to a user. In general, haptic systems work with a 1kHz control loop, that is the forces and hence the currents that are applied to the

actuators are recalculated 1000 times per second according to an internal modelling. For the point based interaction devices such as the phantom, the "correct" feedback force is often calculated from positional data only. If the interaction point (pen tip or thimble tip) is outside a modelled soft object, no forces are applied. Once the point starts to enter into the modelled object, a force corresponding to the distance from the object's surface and the softness of the modeled material is calculated and corresponding currents generated for the actuators. Other properties like friction and viscosity can be modelled accordingly. For other applications, such as a dial with force feedback, the forces can often directly be inferred from a model, as was done with the CharacterPump (see section "AlternativeTextInput" above).

References

Grigore C. Burdea (1999) Keynote Address: Haptic Feedback for Virtual Reality, Proceedings of International Workshop on Virtual prototyping, Laval, France, pp. 87-96, May 1999.

3.6 Taste

The modality 'taste' is for practical reasons largely unexplored in HCI.

3.7 Smell

3.7.1 Output

3.7.1.1 Fragrance generation

Scents constitute an important source of information for people during their daily routines. Scents are said to influence people's mood, their product choice, the time they spend in a shop, the impression they form of others, the attraction towards other people, the food they eat, the things they re-member and it can even warn them for potential danger, such as gas leaks. In addition, people with smelling disabilities often experience depressions (Van Toller & Dodd, 1991; Schaffelaars, 1997), eating disorders and even sexual problems (Falck & Schaffelaars, 1999). The importance of the sense of smell has been largely ignored by the general audience and also by science. The main reasons for this seem to be twofold, on the one hand people are often not consciously aware of the influence or even plain presence of a scent (Köster, 2003) and on the other hand there appear to be negative social attitudes and taboos surrounding smells. Moreover studying scents involves several practical problems, like problems with maintaining a constant perceived odor intensity everywhere in a room, the variability between subjects in their sensitivity and their liking for certain scents, and scent neutralization. These problems are largely due to the undeveloped knowledge in the field. The importance of scents in our daily lives is illustrated by the fact that in 1994 only 20% of the perfume industries' income came from making perfumes to wear and 80% came from perfuming objects in our lives (Barfield & Danas, 1996). Such objects included products like household cleansers, toiletry and decoration material. To give an example of the turnover in this market; in 2003 the sales of air fresheners in the Netherlands was 50 million euros.

Japan's NTT Communications Corp., of Tokyo, is working on the development of its Kaori Tsushin, or Fragrance Communications, as a way to pull our noses into the equation. The telecom and network services company has come up with an Internet-linked fragrance system that can be used to

generate a wide variety of scents on demand with the aim of heightening experiences, influencing moods, and maybe opening wallets (see <http://www.spectrum.ieee.org/jan08/5848>).

Another example is the olfactory game called 'Fragra' which combines visual information together with smells, this can be found on <http://chihara.aist-nara.ac.jp/mtheater/y2004/200402.html>.

There is also research devoted to the problem of emitting smell locally which means local in the sense of delivery at a specified location such as in the face of the user (<http://portal.acm.org/citation.cfm?id=766109>).

References

Barfield, W., Danas, E. (1996). Comments on the use of olfactory displays for virtual environments. *Presence*, 5, 109-121.

Falck, M., Schaffelaars, D. (1999). *Geur en ontwerp*. Eindhoven: ZOO producties.

Köster, E.P. (2003). The psychology of food choice: some often encountered fallacies. *Food Quality and Preference*, 14, 359-373.

Van Toller, S., Dodd, G.H. (1991). *Perfumery: the psychology and biology of fragrance*. London: Chapman & Hall.

3.8 Thermoception - feeling temperature

Thermoception has found little attention in HCI, although in principle it is well suited to transmit subtle information, especially for wearable devices. As an example from another area, there has also been some work on using temperature for more realistic haptic rendering in VR (Dionisio1999).

References

Jose Manuel Simoes Dionisio (1999) "Physically-based thermal feedback for real-time user interaction in Virtual Environments", Hrsg.: Fraunhofer-Institut für Graphische Datenverarbeitung IGD, Darmstadt 1999, ISBN 978-3-8167-5247-9.

3.9 Nociception - feeling pain

The modality 'pain' is for practical reasons largely unexplored in HCI. There is an application of nociception to a videogame installation: <http://www.painstation.de>.

3.10 Equilibrioception -- feeling balance

3.10.1 Input

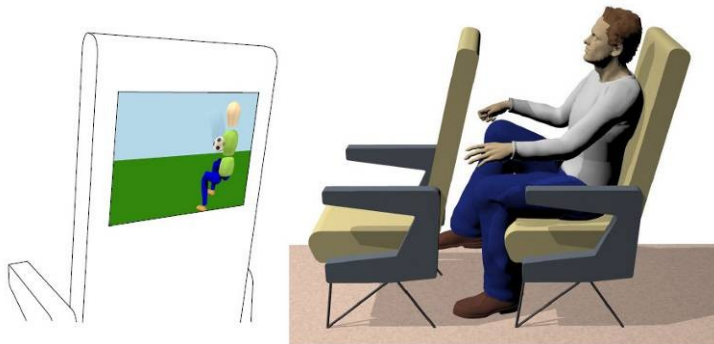
3.10.1.1 Tilting

Besides pointing with a pen or finger, another method to operate especially small portable devices has been proposed: operation by tilting the device as a whole (Rekimoto96). The angle of tilt for a

display that is held horizontally can be easily measured in both the left/right and up/down directions by means of accelerometers. Tilting has become popular also with the iPhone that switches the display automatically between portrait and landscape mode according to whether the device is held upright or horizontally, although there have been comparable applications before.

3.10.1.2 Gesture recognition

In the SEAT project(<http://www.seat-project.org/>), sensors are integrated into airplane seats to detect gestures as input for in-flight entertainment systems, in order to reduce both psychological and physical stress in air travel. The content provided by the entertainment systems helps to reduce psychological stress, and gesture recognition is used as input for the content. The use of this system thus also stimulates passengers in limited seating space to make movements that would result in the reduction of physical stress (Westelaken et al, 2008).



References

Jun Rekimoto (1996) "Tilting Operations for Small Screen Interfaces", User Interface and Software Technology (UIST'96), 1996.

Westelaken, R. van de, Hu, J., Liu, H., Rauterberg, G.W.M. (2008). Integrating gesture recognition in airplane seats for in-flight entertainment. In Z. Pan., X. Zhang, A. El Rhalibi, W. Woo, L. Yi (Eds.), *Technologies for E-Learning and Digital Entertainment ; Third International Conference, Edutainment 2008, Nanjing, China, June 25-27, 2008. (Lecture Notes in Computer Science, Vol. 5093, pp. 353-360)*. Berlin: Springer-Verlag.

4 Sensors and enabling technologies

This Chapter provides an overview of technologies which are not specifically developed to enable user interaction but can be considered to be key enabling technologies for one or more interaction technologies. These technologies are mainly basic sensing and positioning technologies which form the technology background for specific interaction technologies which are described in Chapter 4.

4.1 Ultrasound

Sound, both in audible and ultrasound regions, can be used in a number of ways in location technology. The most common methods use transducers at wavelengths too short to be audible for humans (ultrasound) and time-of-flight measurements to measure distances. Through triangulation, cm range position determination is possible. The transducers are used in one of three major configurations:

1. A fixed infrastructure of transmitters (base stations) with portable receivers on devices or people. This configuration is most privacy secure, because it is the portable device that determines its own position, similar to GPS systems.
2. A fixed infrastructure of receivers with active transmitters on the devices or people. In this case the system tracks the positions. Advantage is a much simpler portable tag than in the case of portable receivers, but in this configuration users or devices are being tracked, so additional measures are needed to secure privacy when needed.
3. Ultrasound reflection measurements, either installed in the environment or on a mobile device. This configuration is typically used to measure proximity. An example is parking sensors in automobiles.

In more advanced embodiments, speed measurements and mapping is possible. The speed of sound is dependent on air temperature, so compensation is needed for high precision measurements. Signals arriving after reflections and multiple reflections can usually be separated easily from the direct line-of-sight signals, but can result in deterioration of update rates in tracking. However, it is also possible to turn this around and make use reflections to create virtual ultrasound sources in a room. In this way the number of base stations needed can be reduced, in principle even to one. [ref: E.O. Dijk, C.H. van Berkel, R.M. Aarts and E.J. van Loenen, A 3D Indoor Positioning Method using a Single Compact Base Station, Proc. PerCom 2004 (2nd IEEE Int. Conf. on Pervasive Computing and Communications), Orlando, 101-110 (2004)].

4.2 Passive IR

A Passive InfraRed sensor (PIR sensor) is an electronic device that measures infrared (IR) light radiating from objects in its field of view. PIR sensors are often used in the construction of PIR-based motion detectors. Apparent motion is detected when an infrared source with one temperature, such as a human, passes in front of an infrared source with another temperature, such as a wall.

All objects emit what is known as black body radiation. It is usually infrared radiation that is invisible to the human eye but can be detected by electronic devices designed for such a purpose.

The term passive in this instance means that the PIR device does not emit an infrared beam but merely passively accepts incoming infrared radiation.

PIR sensors belong to the class of thermal detectors. Thermal detectors can measure incident radiation by means of a change in their temperature. When an appropriate absorbing material is applied to the detector element surface, they can be made responsive over a selected range of wavelengths. PIR sensors are designed to detect human bodies, thus the wavelengths of interest are mainly in the range of the infrared window at 8-14 μ m, in which the IR emission of bodies with a temperature of 37°C also peaks.

The basic structure of a pyroelectric sensing element is a planar capacitor whose charge Q is proportional to the temperature rate and the detector's surface area ($\Delta Q = A \cdot p \cdot \Delta T$ where A is the area of the sensing element and p the pyroelectric coefficient specific for that material). This charge is measured using electrodes as a current through a capacitor surface $I = A \cdot p \cdot \frac{dT}{dt}$. In typical

Commercial Off The Shelf (COTS) PIR sensors the output do not exceed few tenth of μ A, thus it must be amplified with a high impedance preamplifier, typically a FET transistor in source follower configuration is used.

COTS PIR sensors usually include two or four sensitive elements in order to improve the immunity to changes in the background temperature and achieve a shorter settling time. The resulting schematic is presented in figure -PIR1-.

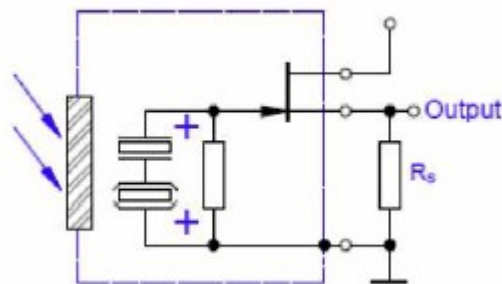


Figure 2: Schematic of a typical PIR sensor

PIR sensors are used in conjunction with an array fresnel lenses that are used both to shape the Field Of View (FOV) of the sensor and to modulate incident radiation. In fact, since the sensor is sensitive to changes in input infrared radiation, the array produces in a series of distinct cones of view. As the body moves within the FOV it moves in and out different cones producing an observable output.

Being passive (thus, low-power), low cost and presenting a small form factor (see figure –PIR2-), PIR sensors are well suited for application for Wireless Sensor Network (WSN).

PIR sensor are widely used to in surveillance systems (Moghavvemi, 2004) and automatic light switching systems (see figure –PIR2-).



Figure 2. PIR sensor typical applications

Recent work used PIR sensor for people tracking. In (Zappi, 2008) an array of PIR sensor is used to detect people position within an hallway. In (Gopinathan, 2003) a pyroelectric motion tracking system able to detect the path of a single person moving in an area and based on coded apertures has been developed. The apertures are designed to modulate the visibility of four PIR detectors over a 1.6×1.6 m area such that the position of a source among 15 resolution cells may be discriminated using 4 elements.

In (Song, 2008) the performance and the applicability of the PIR sensors for security systems and propose a region-based human tracking algorithm is analyzed. Results shows that the human tracking algorithm based on the PIR sensors performs very well with the proposed sensor deployment. Slightly different is the work presented in (Hashimoto, 1997) where an array of PIR is used to count the number of people moving through a gate. Since the sensor can only detect temperature changes, the incident radiation flow is modulated by a chopper wheel that temporarily obstruct the PIR Field Of View (FOV). The data from the sensors is processed by a PC.

Sensor networks implemented with PIR are useful where privacy must be preserved together with security. In (Rajgarhia, 2004) cameras and PIR sensors are deployed respectively in public and private areas, and their information combined to correlate events such as tracking human motion and undesired access or presence in private areas, such as theft. This work demonstrate benefits of reducing camera deployment in favor of PIR sensors and reports results from a survey on 60 people, stating that people consider motion sensors less invasive for their privacy than cameras.

PIR sensors are often combined with vision systems and other kind of sensors in research focused on robot navigation and localization. In (Sekmen, 2002) a sound source localizer and a motion detector system are implemented on a human service robot called ISAC, with the purpose of redirect the attention of ISAC. The motion detector system use an infrared sensor array of five PIR sensors and it is integrated with the vision system of ISAC to perform realtime human tracking, in an inexpensive way.

Combining PIR sensor with video systems is a common approach to prove video analysis. In (Bryant, 2003) PIR sensors are used to provide a trigger event in a motion-detection application based on cameras for tracking events at night. The appearance of an infrared radiating body set off the PIR sensor, which turns on a floodlight enabling the cameras to capture clearly an event such as animals passing by an outdoor detected area. In (Bai, 2008) a system based on an ARM board with a Camera module being triggered by a Pyroelectric Infrared Sensor (PIR) which senses changes in

the external temperature from an intruder is presented. The system captures the relevant images and send them to a remote server. Finally PIR sensor can be used to improve camera based localization (Cornel, 2007) or tracking (Cucchiara, 2006)

References

- Bai Y.-W. and Teng H. (2008). Enhancement of the sensing distance of an embedded surveillance system with video streaming recording triggered by an infrared sensor circuit. *SICE Annual Conference*, pages 1657–1662.
- Bryant P. and Braun H. W. (2003). Some applications of a motion detecting camera in remote environments. *Technical report*, University of California San Diego.
- Cornel B.(2007). Recognition system using video and infrared information fusion. *Computational International Symposium on Intelligence and Intelligent Informatics*. ISCIII '07., pages 281–283.
- Cucchiara R., Prati A., Vezzani R., Benini L., Farella E. and Zappi P. (2006). Using a wireless sensor network to enhance video surveillance. *Journal of Ubiquitous Computing and Intelligence (JUCI)*, 1:1–11.
- Gopinathan U., Brady D. and Pitsianis N.(2003). Coded apertures for efficient pyroelectric motion tracking. *Opt. Express*, 11(18):2142–2152.
- Hashimoto K., Morinaka K., Yoshiike N., Kawaguchi C. and Matsueda S.(1997). People count system using multi-sensing application. *International Conference on Solid State Sensors and Actuators*.
- Moghavvemi M. and Seng L. C. (2004). Pyroelectric infrared sensor for intruder detection. *TENCON 2004. IEEE Region 10 Conference*, pp 656–659 Vol. 4.
- Rajgarhia A., Stann F. and Heidemann J.(2004). Privacy-sensitive monitoring with a mix of IR sensors and cameras. *In Proceedings of the Second International Workshop on Sensor and Actor Network Protocols and Applications*, pages 21–29, Boston, Massachusetts, USA.
- Sekmen A., Wilkes M. and Kawamura K. (2002). An application of passive humanrobot interaction: human tracking based on attention distraction. *IEEE Transactions on Systems, Man and Cybernetics, Part A*, 32(2):248–259.
- Song B., Choi H. and Lee H. S.(2008). Surveillance tracking system using passive infrared motion sensors in wireless sensor network. *International Conference on Information Networking*, pag 1–5.
- Whatmore R. (1986). Pyroelectric devices and materials. *Reports on Progress in Physics*, vol. 49, no. 12, pp. 1335–1386.
- Zappi P., Farella E. and Benini L. (2008). Pyroelectric infrared sensors based distance estimation. *Sensor 2008*, Lecce.

4.3 Active IR and laser scanning

Proximity of objects can be detected using reflection of visible or (preferably) invisible light. Proximity sensors based on IR LED's are low cost and easy to implement.

High precision mapping and tracking of objects over larger distances and areas can be accomplished using laser scanning. This technique is relatively expensive, and mostly used in industrial applications.

4.4 GPS

GPS devices use RF signals transmitted by a set of satellites with accurately known positions w.r.t. the globe, to measure their distances and through triangulation the position of the GPS receiver. High update rates also allow speed and direction measurements. Inaccurate when used indoors due to occlusion and reflections, but indoor GPS has been demonstrated using a dedicated indoor transmitter infrastructure.

4.5 Head movement trackers

Head mounted displays used in virtual reality applications are often equipped with head trackers. These movement sensors are often based on mechanical principles (acceleration sensors). Computer vision technology provides an alternative approach to sense head movement.

4.6 Accelerometers

An accelerometer is a device for measuring acceleration and gravity induced reaction forces. Single- and multi-axis models are available to detect magnitude and direction of the acceleration as a vector quantity. Accelerometers can be used to sense inclination, vibration, and shock. They are increasingly present in portable electronic devices and video game controllers. They can be applied for motion detection and gesture recognition in cases where tagging is acceptable.

Accelerometers are sensing transducers that provide an output proportional to their acceleration. The most popular class of accelerometers is the piezoelectric accelerometer. This type of sensor is capable of measuring a wide range of dynamic events. However other technologies are used such as piezoresistive, capacitive and servo.

Piezoelectric materials produce an electrical signal proportional to the mechanical stress applied to them. A different technique is emerging in the last years thanks to the evolution in the field of microelectronics and is based on Micro Electro Mechanical System (MEMS). Typical MEMS accelerometer is composed of movable proof mass with plates that is attached through a mechanical suspension system to a reference frame. Movable plates and fixed outer plates represent capacitors. The deflection of proof mass is measured using the capacitance difference. Acceleration is then derived from this displacement once knowing the mass of the moving plate.

Accelerometers applications are countless.

They are used in industrial application to predict failures on machines (Chindurza 2004), motors

(Jagannath, 2007) and, in general, rotating equipment (Nandi 1995). The typical approach is to analyze the spectrum of vibration of operating equipment and to detect changes from a reference signature (Rocha, 2005 and Cempel, 2003).

Accelerometers are used in automotive for airbag systems (Mahmud, 1995), to estimate real-time the tire-road friction coefficient (Rajamani, 2006) and misfire detection (Villarino, 2004).

Application of accelerometers in low cost consumer business include hard disk control (Jinzenji, 2001) and mobile phones (Apple, 2007).

Accelerometers can be used in health care application for rehabilitation (Rocchi, 2008), elderly care (Noury, 2002, and Scanaill 2006) and remote health care (Jizi, 2005).

Thanks to their reduced size often they are embedded into smart objects for video game controlling (Nintendo, 2008), or tangible interfaces (Baraldi, 2007).

Accelerometers are used in gesture recognition as an alternative to cameras. In (Wan, 2008) the design and implementation of a HCI using a small hand-worn wireless module with a 3-axis accelerometer as the motion sensor is presented. The vision of a world where a large number of unobtrusive sensors are embedded into user garments to recognize activities is tackled in (Zappi, 2007) where a hierarchical approach for gesture recognition is presented.

References

Apple (2007), <http://www.apple.com/iphone/features/accelerometer.html>

Baraldi S., Benini L., Cafini O., Del Bimbo A., Farella E., Landucci L., Pieracci A., Torpei N., (2007) Introducing TANGerINE: A Tangible Interactive Natural Environment. *In proceedings of ACM MultiMedia 2007*, Augsburg, 24-29.

Cempel C. (2003). Multifault condition monitoring of mechanical systems in operation. *In Proc. IMEKO XVII*, Croatia, pp. 1–4.

Chindurza, I., Dorrell, D.G. and Cossar, C. (2004). Vibration analysis of a switched-reluctance machine with eccentric rotor. *Power Electronics, Machines and Drives, 2004. (PEMD 2004). Second International Conference on (Conf. Publ. No. 498)*, vol.2, no., pp. 481-486 Vol.2, 31

Jagannath, V.M.D., Raman, B. (2007). WiBeaM:Wireless Bearing Monitoring System. *2nd International Conference on Communication Systems Software and Middleware 2007.*, vol., no., pp.1-8, 7-12.

Jinzenji A., Sasamoto T., Aikawa K., Yoshida S. and Aruga, K.(2001). Acceleration feedforward control against rotational disturbance in hard disk drives. *Magnetics, IEEE Transactions on*, vol.37, no.2, pp.888-893.

Jizi Li, Chunling Liu (2005). An ambulatory monitoring architecture for the remote healthcare service system. *Services Systems and Services Management, 2005. Proceedings of ICSSSM '05. 2005 International Conference on* pp. 1414-1418 Vol. 2.

Luinge, H.J., Veltink, P.H.(2004) Inclination measurement of human movement using a 3-D accelerometer with autocalibration. *Neural Systems and Rehabilitation Engineering, IEEE*

Transactions on , vol.12, no.1, pp.112-121.

Mahmud, S.M. and Alrabady, A.I.(1995). A new decision making algorithm for airbag control. *Vehicular Technology, IEEE Transactions on* , vol.44, no.3, pp.690-697.

Nandi, A.K., Dickie, J.R., Smith, J.A. and Tutschku, K.(1995). Classification of conditions of rotating machines using higher order statistics. *Higher Order Statistics in Signal Processing: Are They of Any Use? IEE Colloquium on* , pp.3/1-3/6.

Nintendo (2008), <http://wii.com/>

Noury N. (2002).A smart sensor for the remote follow up of activity and fall detection of the elderly. *Microtechnologies in Medicine & Biology 2nd Annual International IEEE-EMB Special Topic Conference on* , pp.314-317.

Rajamani R., Piyabongkarn D., Lew J.Y. and Grogg J.A.(2006).Algorithms for real-time estimation of individual wheel tire-road friction coefficients. *American Control Conference* , pp.14-16.

Rocchi L., Benocci M., Farella E., Benini L., Chiari L. (2008) Validation of a Wireless Portable Biofeedback System for Balance Control: Preliminary Results. *Pervasive Health Conference*. Tampere Finland.

Rocha, L.A., Cretu, E., Wolffenbuttel, R.F. (2005). MEMS-based mechanical spectrum analyzer. *Instrumentation and Measurement, IEEE Transactions on* , vol.54, no.3, pp. 1260-1265.

Scanail C.N., Ahearne, B. and Lyons, G.M. (2006). Long-term telemonitoring of mobility trends of elderly people using SMS messaging. *Information Technology in Biomedicine, IEEE Transactions on* , vol.10, no.2, pp.412-413.

Villarino R. and Bohme J.F. (2004). Pressure reconstruction and misfire detection from multichannel structure-borne sound. *Acoustics, Speech, and Signal Processing. (ICASSP '04). IEEE International Conference on* , vol.2, pp. 17-21

Wan, Silas, Nguyen, Hung T. (2008). Human computer interaction using hand gesture. *30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2008.*, vol., no., pp.2357-2360, 20-25

Zappi P., Stiefmeier T., Farella E., Roggen D., Benini L., Tröster G.(2007). Activity recognition from on-body sensors by classifier fusion: sensor scalability and robustness. *in Proceedings of The Third International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP) 2007*, pages 281-286, Melbourne.

4.7 Pressure sensors, load cells

Pressure sensors can be used in various configurations, from single sensors in for example car seats to determine if the seats are occupied to high resolution pressure sensing mats that can be used to measure presence, position, orientation, weight and even identity of people on a floor. Other powerful embodiments are pressure sensors mounted under the legs of for example chairs, tables or floor tiles. By careful analysis of the changes in pressure patterns, information on location of people and objects can be obtained. In this way it is for example possible to detect the location of a coffee cup on a table, and through total weight measurement also the activity of drinking from a cup, etc.

4.8 Environmental sensors

Sensors such as temperature sensors, humidity sensors, or light sensors can be used to measure environmental changes, which can in turn be related to activity patterns.

4.9 Contact sensors

Simple contact sensors can already provide valuable information on activities of people, such as holding a steering wheel or pushing a shopping cart. They can also be used in positioning: when touching of objects with a known position is detected, the position of the person touching the object is also known.

4.10 Device usage logging

Information on activities can in some cases simply be obtained by logging the usage of devices such as keyboards, mobile phones etc. For example, if typing of a longer text on a fixed keyboard is detected, it is likely that the person typing is seated at the desk.

4.11 Physiological measurements

Physiological measurements such as EDA (electrodermal activity, also known as GSR or galvanic skin response), ECG (electrocardiogram) and EEG (electroencephalogram) may give an indication of a person's activity by means of measured arousal (EDA), heart rate (ECG). In addition, breathing patterns can be measured.

4.12 RF-ID

Radio-frequency identification (RFID) is an automatic identification method, relying on storing and remotely retrieving data using devices called RFID tags or transponders. The technology requires cooperation of an RFID reader and an RFID tag.

An RFID tag is an object that can be applied to or incorporated into a product, animal, or person for the purpose of identification and discrete tracking. The tag is battery free, and uses the magnetic field generated by the reader to briefly power the IC inside, detect its presence, and retrieve its ID. Other types of tags are actively transmitting, using a tiny on board battery, with the advantage that they can be read up to about 10m away and beyond the line of sight of the reader.

In the past, RFID technology has become one of the most applied ubiquitous computing technologies (Langheinrich (2007)). RFID has several advantages over other more traditional identification systems:

- an increased level of automation (e.g. in comparison with bar codes techniques);
- a high information density for identification of goods and/or its features;
- a lot of integration possibilities because of its wireless character.

First and foremost, RFID is applied for logistic processes:

- Pfizer is planning to apply RFID for securing pharmaceuticals against imitation;
- BMW, Airbus Industries, and Maxdata are using RFID for assembling processes;
- Airports use RFID for luggage handling

- The Deutsche Bahn AG uses RFID for controlling railway services.

But more and more RFID is used for end-consumer business and in-store applications. For example, the retailer Galeria Kaufhof in Germany has developed an intelligent fitting room that is able to identify clothes and to give some advice on price, manufacturer, care, or other products (VDI nachrichten, 2008). Also, RFID equipped everyday objects were presented, such as a trolley for health assistance (Hellenschmidt, 2007), or RFID equipped personal souvenirs for getting access to photos (van den Hoven, 2005), or intelligent bottles of wine (Wahlster et al., 2008). Technically problematic is still the identification of goods that have a high share of water (due to the usage of micro waves); thus the itemization of all goods in a trolley is not possible now. Till now only 50 per cent decision makers are convinced that RFID will bring in some benefit into business processes (Isermann, Glaser and Neubert, 2008). Reason for that are still missing technical reliabilities, but also demonstrative applications that bring some added value against still established technologies. One major obstacle is the missing standards (no RFID-off-the-shelf), so that each application has to be planned from the scratch. This is due to a large variety of different RFID products, each of them only appropriate for a limited amount of different applications. At present there are some public fears of RFID's alleged capability for comprehensive surveillance that have prompted a flurry of research trying to alleviate such concerns. Here, for instance, the US state Washington enacted a law that forbids reading personalized RFID tags without the owner's permission. Also the German Federal Office for Information Security (<http://www.bsi.bund.de>) presented technical guidelines for RFID that should ensure the protection of personal data.

References

- Langheinrich M. (2007). *RFID and Privacy*, in: Milan Petkovic, Willem Jonker (Eds.): Security, Privacy, and Trust in Modern Data Management. Springer, ISBN 978-3-540-69860-9, pp. 433-450, Berlin Heidelberg New York, July, 2007.
- VDI nachrichten, *RFID durchdringt interne Logistikprozesse; Den Koffern das Denken beibringen; RFID gibt es nicht von der Stange; RFID-Chips erleichtern E-Loks den grenzüberschreitenden Verkehr*, VDI nachrichten, Hannover, 2008.
- Hellenschmidt M., and Kamieth F. (2007). *BERNIE – Consultant for Nutrition and Intelligent Shopping*, Workshop Proceedings of AmI-07, Darmstadt, November, 2007.
- Wahlster W., Kröner A., Schneider M., and Baus J. (2008). *Sharing Memories of Smart Products and their Consumers in Instrumented Environments*, *it-Information Technology*, 50, (1), 45-49, Oldenbourg, München, Germany, 2008.
- Van den Hoven E., and Eggen B. (2005). *Personal souvenirs as Ambient Intelligent Objects*, Proceedings of the 2005 joint Conference on Smart objects and Ambient Intelligence, pp.123–128, Grenoble, France, 2005.
- Isermann M., Glaser U., and Neubert V. (2008). *Potenzial von RFID in deutschen Unternehmen weitgehend ungenutzt*, Fraunhofer-IPT, 2008.

4.13 Blind object identification

Blind object identification is strictly related to Ambient Interaction paradigm via remote wireless technologies, such as the RFID technology which is based on two fundamental sets of wireless devices: 1) a number of tags, which consists of electronic chips coupled with an antenna, distributed in the ambient. Tags may be placed in non visible sites attached to artifacts and store data about the objects which need to be identified. 2) a reader, which can be both a standalone device and integrated in a smart phone carried by the user and combined with an external antenna, reads/writes data from/to a tag via radio frequency and transfers data to a host computer.

An enormous number of applications, has been already demonstrated in the field of building construction (Wing, 2006) and non-destructive structural health monitoring (Rizzoli et al. 2009) in order to maintain awareness on dynamic construction site, improve knowledge of the location and movement of workers, construction equipment, and manufactured construction components. Automatic determination of information from intelligent site is then available due to feedback from sensors that could provide construction object tracking.

Furthermore RFID technology allows to uniquely identify facility components, store their maintenance history on site and get informations on demand. Recent studies and applications tested the performance of active ultra high frequency RFID technology on facility components during operations and maintenance phase repetitively over an extended period of time (Wing, 2006). The physical space (e.g. a smart room) where physical objects (i.e. RFID-enabled objects) can be selected in order to activate associated services then acts as an extension to the User Interface of a smart device where new applications are activated in the same way as clicking on an icon on a conventional display. Discovery and interaction do not need specific knowledge of local network configuration: object may be selected, clicked upon or dropped on different applications depending on the user's intentions. As a result, different actions may be invoked. RFID-enabled smart entities are able to self-configure when they first join the network, share their presence announcement through the RFID interface, automatically update it while connected to the network and exit gracefully (Antoniou et al. 2006, Chumkamon et al. 2008).

In spite of these benefits, RFID technology comes with some limitations. Reading/writing ranges depend on the operation frequency low, high, ultrahigh, and microwave and whether the tag needs a battery to operate or not. Active tags have typically higher reading ranges; however, they have a limited lifetime requiring battery replacement periodically. For passive tags, reading range and memory are limited. The ambient is usually harsh (Dobkin 2005, Wagner 2007, Engels 2002) from the radio propagation point of view, due to the specific layout of the smart space and choice of the optimum location and device/operating frequency selection need be accurately designed (Rizzoli et al. Sept. 2006, Dec. 2006).

References

R. Wing, RFID applications in construction and facilities management, *Electronic Journal of*

Information Technology in Construction 11 (2006) 711–721.

V. Rizzoli, A. Costanzo, E. Montanari and Andrea Benedetti, "A New Wireless Displacement Sensor Based On Reverse Design Of Microwave And Millimeter-Wave Antenna Array", *accepted for Publication in the IEEE Sensors Journal in the special issue of Sensor systems for structural health monitoring, expected May 2009*.

[Esin Ergen](#), [Burcu Akinci](#) [Bill East](#), and [Jeff Kirby](#), "Tracking Components and Maintenance History within a Facility Utilizing Radio Frequency Identification Technology"; *J. Comp. in Civ. Engrg.* Volume 21, Issue 1, pp. 11-20 (January/February 2007).

Antoniou, Z.; Krishnamurthi, G.; Reynolds, F.; "Intuitive Service Discovery in RFID-enhanced networks", *Communication System Software and Middleware, 2006. Comsware 2006. First International Conference on*, Page(s):1 –5.

Chumkamon, S.; Tuvaphanthaphiphat, P.; Keeratiwintakorn, P.; "A blind navigation system using RFID for indoor environments", *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, 2008. ECTI-CON 2008. 5th International Conference on*, Volume 2, 14-17 May 2008 Page(s):765 – 768.

Dobkin, D.M.; Weigand, S.M.; "Environmental effects on RFID tag antennas"; *Microwave Symposium Digest, 2005 IEEE MTT-S International*, 12-17 June 2005 Page(s):4 pp
Wagner, J.; Fischer, R.; Gunthner, W.A.; "The Influence of Metal Environment on the Performance of UHF Smart Labels in Theory, Experimental Series and Practice", *RFID Eurasia, 2007 1st Annual*, 5-6 Sept. 2007 Page(s):1 – 6.

Engels, D.W.; Sarma, S.E.; "The reader collision problem"; *Systems, Man and Cybernetics, 2002 IEEE International Conference on*; Volume 3, 6-9 Oct. 2002 Page(s):6 pp. vol.3.

V. Rizzoli, A. Costanzo, D. Masotti, P. Spadoni, and A. Neri, "Prediction of the End-to-End Performance of a Microwave/RF Link by means of Nonlinear/Electromagnetic Co-Simulation ", *IEEE Transactions on Microwave Theory and Techniques*, Vol. 54, No. 12, Dec. 2006, pp. 4149-4160.

V. Rizzoli, A. Costanzo, M. Rubini, and D. Masotti, "Investigation Of Interactions Between Passive RFid Tags By Means Of Nonlinear/EM Co-Simulation", *Proceedings of the 36th European Microwave Conference* (Manchester), Sept. 2006, pp. 722-725.