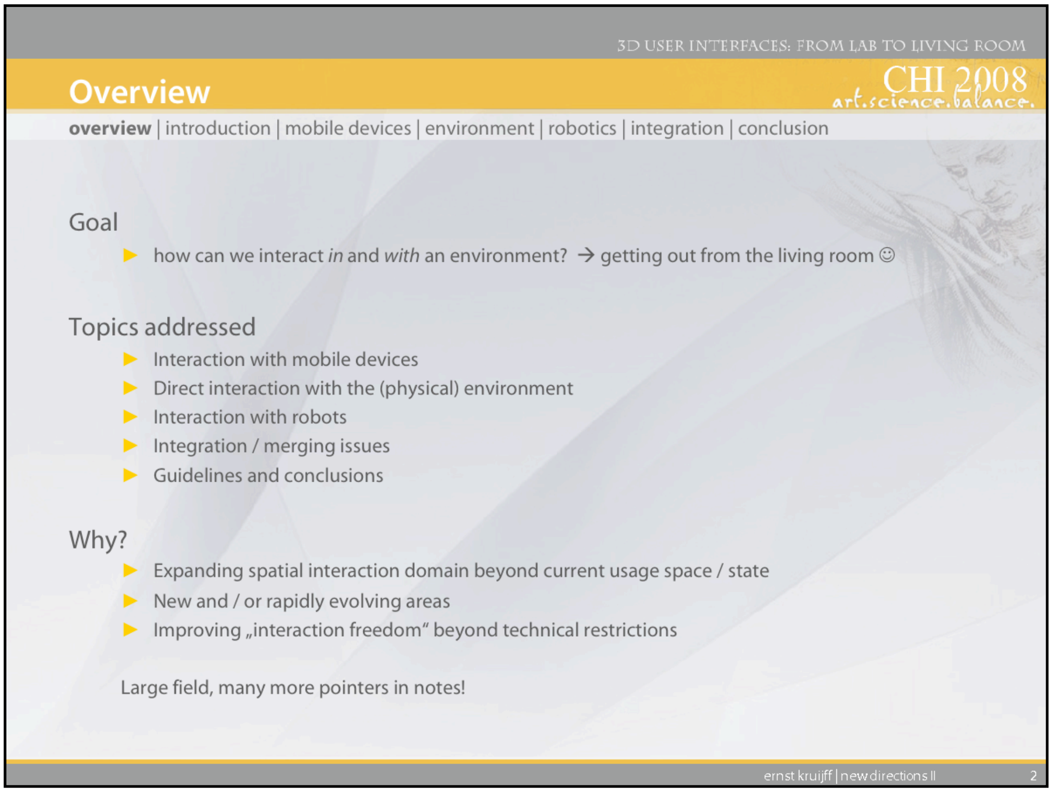
CHI 2008
art.science.balance.

New Directions II

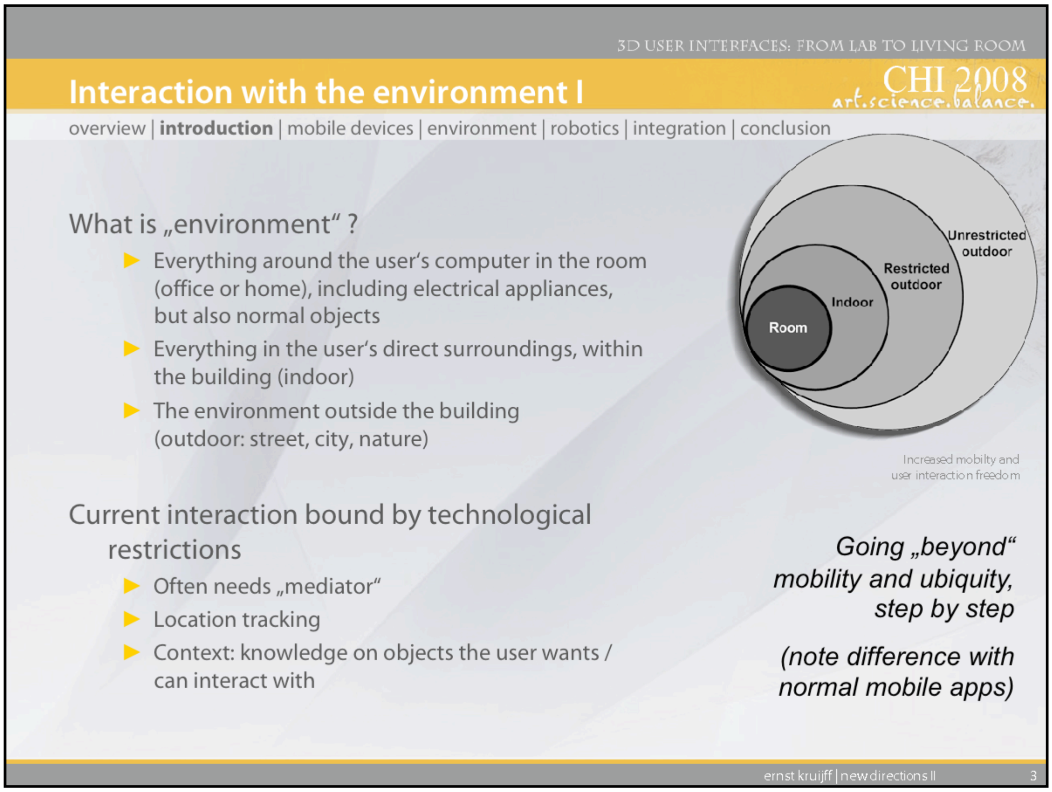
Ernst Kruijff



In this part of the course, we take a look at how we can interact *in* and *with* the „environment around us“, moving away from classical desktop and indoor applications. We focus on interaction using mobile devices, interaction with the environment itself using large scale embedded display systems, and human-robot interaction. All of these fields gain increasing interest, with many open research issues.

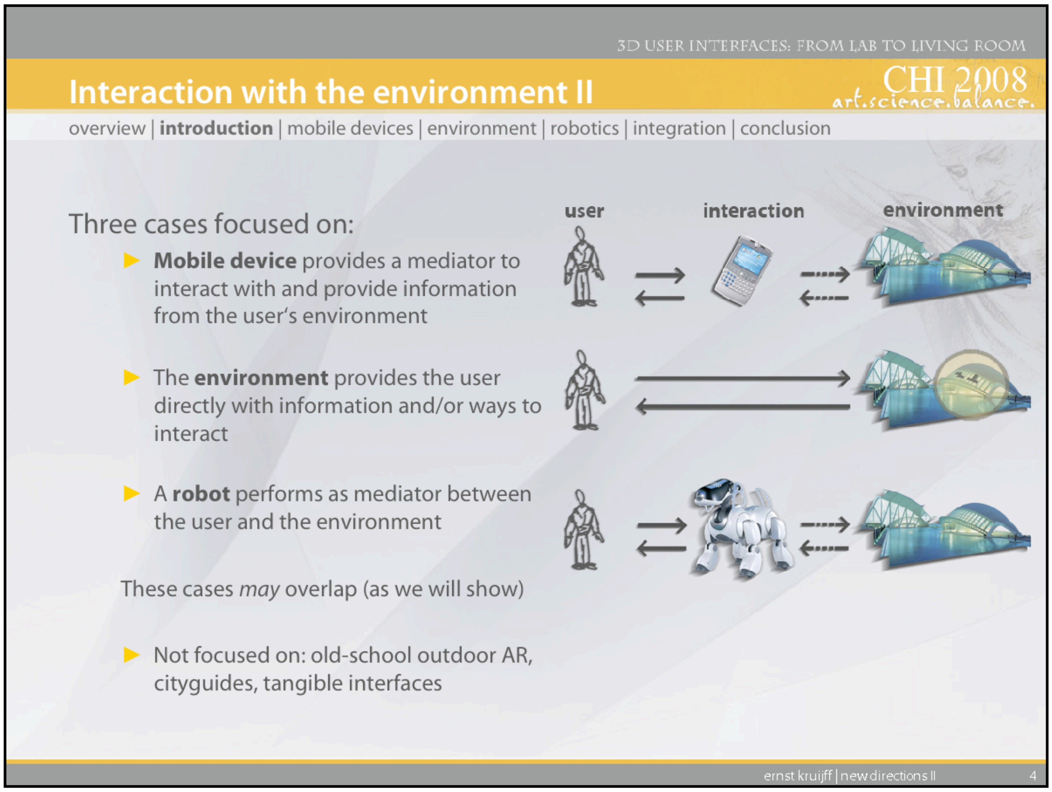
With respect to spatial interaction, all fields pose considerable challenges: in most cases, „traditional“ 3D user interface techniques cannot be used but need to be adapted considerably, or need to be completely re-created. We will provide a general picture of the design issues and challenges, and give some useful examples.

Please note that the field we try to cover is big: Due to scope of the topics and the limits of this talk (time), it is highly recommended to read the literature referenced / proposed throughout the slides.



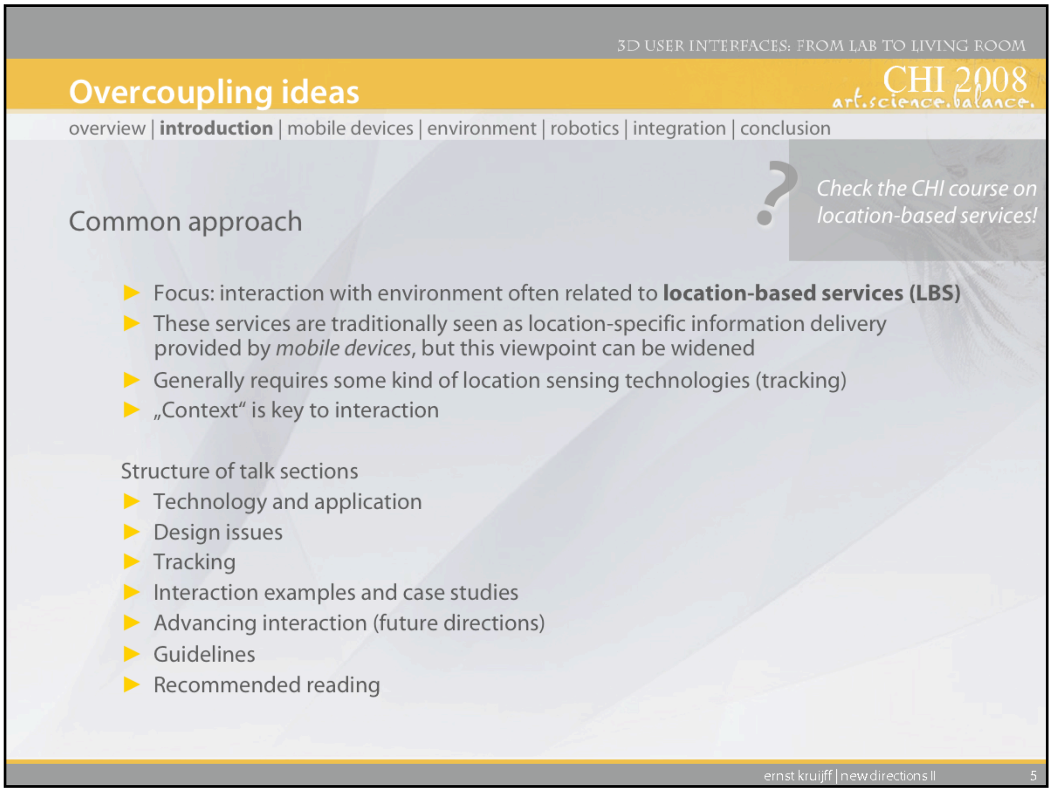
Looking at how users can interact with the environment, we may see many relations to mobile and ubiquitous computing, as introduced by people like Weiser or Mann. We are moving away from predominantly desktop based applications, to situations where we process information „on demand“, both unrelated and related to the location we are. The latter is of high interest to spatial interaction, but also poses great limitations on the user. Using the current state of technology, it is hardly possible to „just“ interact with any object around us. We are dependant on location tracking mechanisms that tell us where we are, or need to have some kind of intelligence stored in the object we interact with. Basically we need to know „where we can perform which tasks with what object“.

Researchers are creating user freedom step by step through such technologies like GPS embedded in phones, but most interaction is still limited. The further we move into the outdoor world, the more restricted interaction gets. In most cases we need some kind of mediator to interact with the environment – these mediators are the key in this talk.



Interaction with the environment often takes place via a mediator: some kind of device that interprets and „commands" the environment around us. Within this talk, we deal with two different kinds of mediators: handheld, mobile devices and robots. Furthermore, we take a look at how the environment itself can function as device itself, by focusing on extremely large embedded display systems. Whereas handled seperately, these ways of interaction can overlap, as we will illuminate at the end of the talk.

Outside the focus of the talk are old-school augmented reality setups (mounted laptops and head-mounted displays), traditional city audio guides, and tangible interfaces. The latter will be handled seperately in the Beyond Visual: shape, haptics and actuation in 3DUI talk.



Interaction with the environment often relates to location-based services (LBS, Gartner et al). Whereas general mobile computing applications are not necessarily location-based, most of the spatial interaction handled in this talk does have a direct reference to a specific location. Whereas LBS,mostly deals with mobile devices, we believe this view can be widened. Location-based interaction with information does not necessarily require a mobile device as mediator, as we show in the sections on environmental displays and robotics.

Location-based services can make use of either push or pull services: pull services deliver information requested by the user, whereas with push services, information is provided that is indirectly or even not at all requested by the user. Location-based services are generally used for information provision that is related to orientation/localisation tasks, or for identification of objects or users in the vicinity. However, one can also envision services that are related to some form of entertainment (infotainment), billing or emergency handling. The key to location-based services is context: the user's location (spatial context), other users in the surroundings (social contaxt) and available / manipulatable content (information context) define the interaction space of the user (Schilit et al ´94).

References

Gartner,Cartwright, Peterson. Location Based Services and TeleCartography. Springer Verlag 2007.

Schilit, B., Adams, N. and, Want, R., 1994. Context-aware computing applications. In: Proceedings of IEEE Workshop on Mobile Computing Systems and Applications. 1994, Santa Cruz, California. , 85- 90.

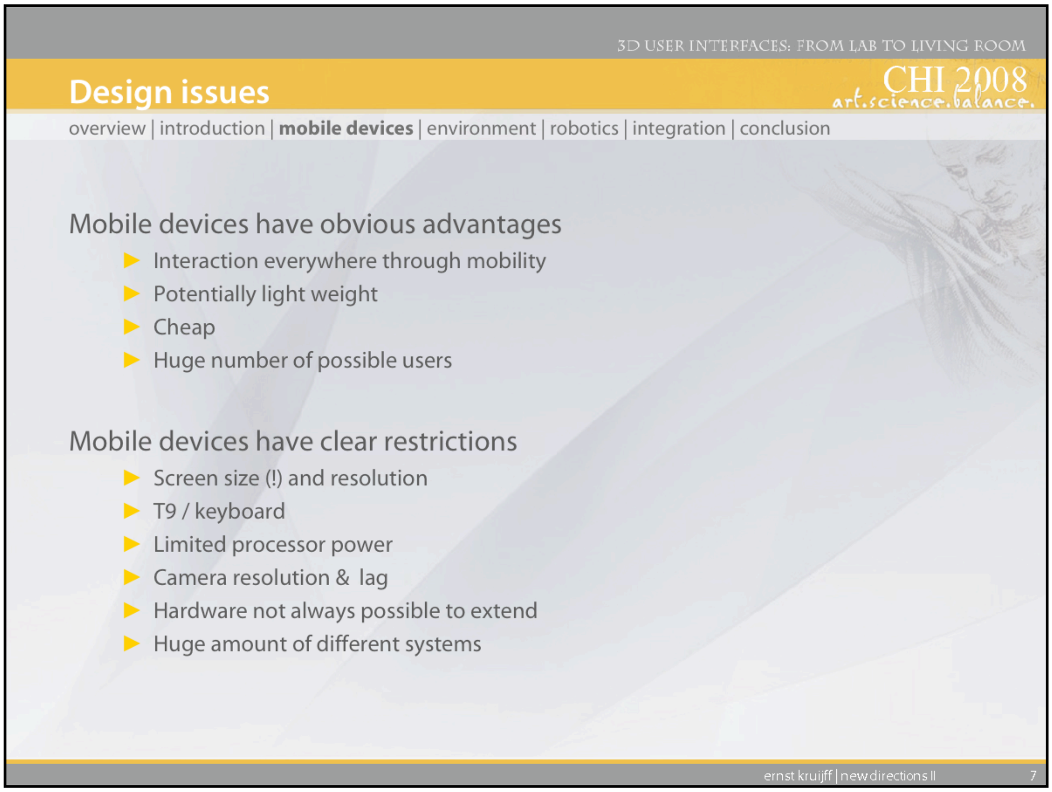
Mobile devices are the first „logical“ choice for interaction in and possibly with the environment us. Whereas most devices are trimmed for desktop-like interaction, most of the devices can potentially enable spatial interaction. Whereas mobile spatial interaction (MSI) used to only be possible with wearable constructions like backpack-based augmented reality setups (like Piekarski&Thomas01), nowadays handheld systems exist that make these older setups obsolete, or at least allow for different kinds of truly portable ways of MSI. PDAs and cell phones have considerable processing capabilities that allow to run simple programs, mostly using some kind of computer-vision based tracking method by interpreting the camera stream from the built in camera. In addition, mobile computer platforms like the ultramobile PC (UMPC) provide enough capacities to run more complex application, and are still quite wearable (around 500 grams).

Application areas are very much similar to the previously mentioned LBS examples. Moreover, it is sometimes difficult to draw the border between mainstream mobile interaction techniques (Jones&Marsden`06) and mobile spatial interaction: many of the basic techniques (such as localization) are the same, and some of the methods used for interaction can be both used for 2D and 3D spaces.

References

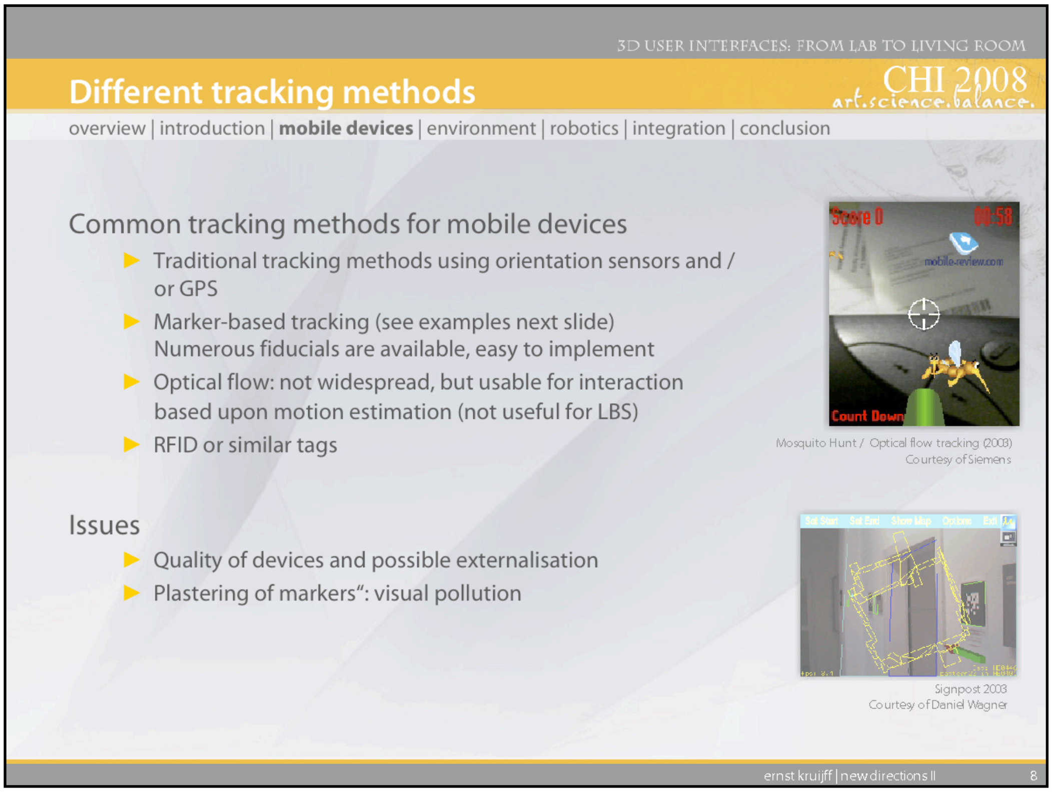
Jones, M. and G. Marsden, Mobile interaction design. 2006: John Wiley & Sons Ltd.

Piekarski, W. and B. Thomas. Tinmith evo5 - An Architecture for Supporting Mobile Augmented Reality Environments. In In Proc. of 2nd International Symposium on Augmented Reality (ISAR'01). 2001.



Mobile devices have considerable advantages that can truly accelerate popularity of applications that make use of MSI. One can interact almost everywhere, especially since the devices are so light. They are cheap and there is a truly huge user base,allowing for new fields like massive multi-user mobile applications to come into reach.

On the other hand, mobile devices can seriously limit user interaction. The screen size can be truly limiting and puts specific constraints on both the interaction space (screen estate) and the amount of attention the user needs to spent on interaction. Additionally, most devices come with terrible control structures for interaction: T9 / keyboard and even most mini joysticks are not good at all for longer interaction, or for interaction at all. Processor and camera capacities are usable, but also limit the complexity of applications and the locations users can interact with content: bad light conditions can easily destroy tracking performance of computer vision-based techniques. Extending devices with additional or better hardware is rather hard most of the times: whereas well possible with a UMPC, most cellphones and PDAs can hardly be extended with hardware.



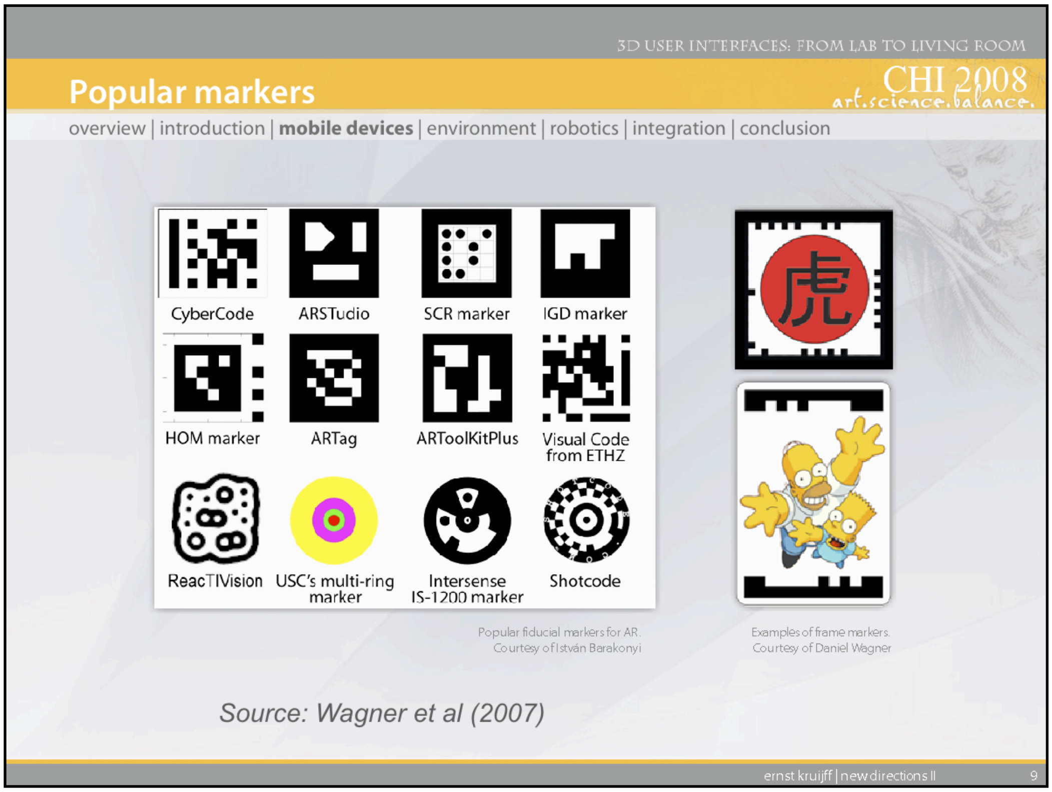
In order to locate mobile devices, a whole range of tracking mechanisms can be used. They include using inertia/orientation sensors and GPS, marker-based tracking systems, techniques that are using optical flow are used, and RFID. Especially the latter is hardly useful for viewpoint dependant spatial interaction, since it provides only very rough position information.

The quality of tracking mostly depends on the sensors being used: most embedded devices (cameras, GPS, etc.) deliver reasonable but not very high quality data. Often, one needs to depend on external solutions that are hard to connect. Moreover, many of the techniques depend on an external source, may it be a marker or a satellite signal. This signal might not be available, or, may lead to unwanted side effects to make it available: in order for marker-based tracking to work, it may need huge amounts of markers that will truly pollute out environment.

References

Christian Reimann, Volker Paelke. Computer Vision based Interaction Techniques for mobile Augmented Reality. In: 5th Paderborn Workshop Augmented and Virtual Reality in der Produktentstehung. HNI-Verlagsschriftenreihe, Nr. 188, Paderborn, Germany.

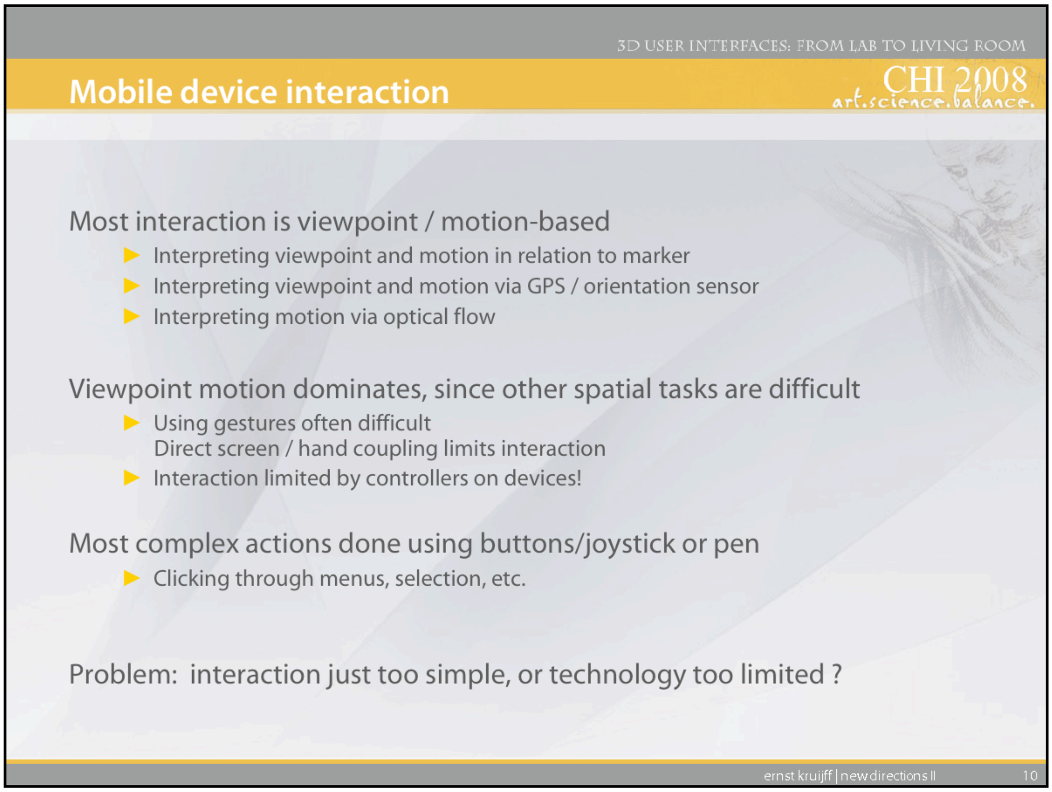
Schmalstieg, Wagner. Experiences with Handheld Augmented Reality. ISMAR 2007.



Some example of fidicial markers, taken from (Wagner et al 07)

Reference

Wagner, Barakonyi, Billinghurst, MacIntyre. Designing and Developing Handheld Augmented Reality Systems. IEEE ISMAR 2007.

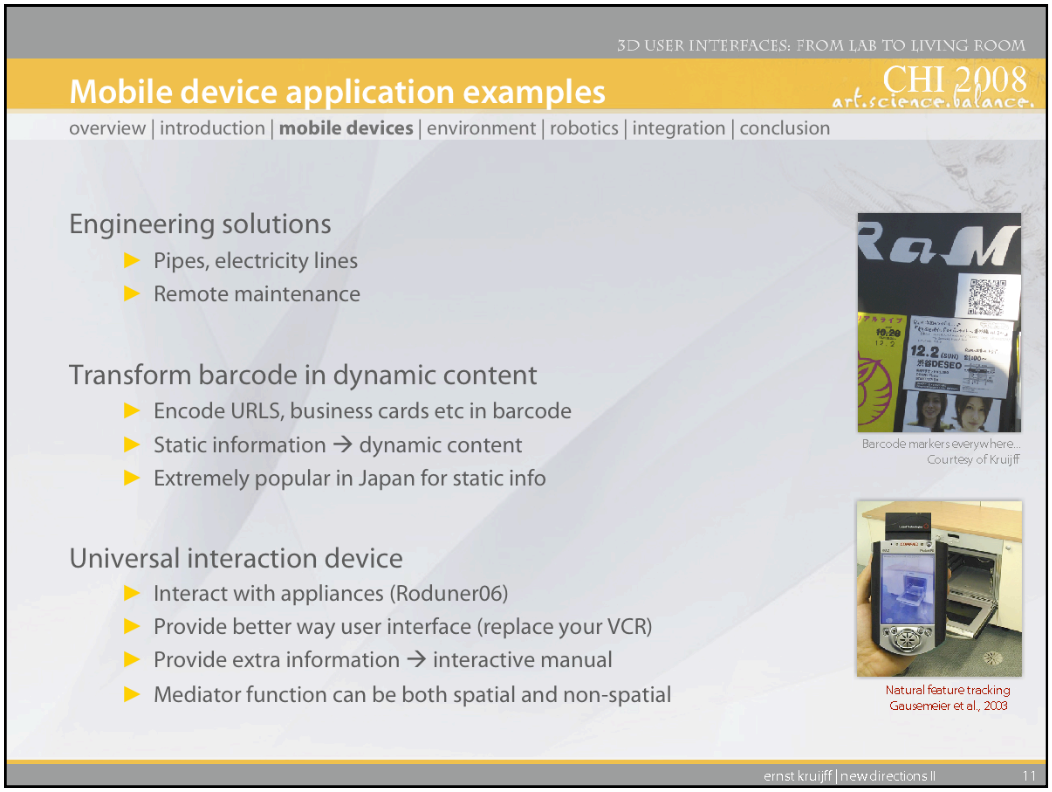


Mobile interaction predominantly is viewpoint / motion-based, by interpreting the sensor data at hand (GPS, video, etc.). Other kinds of spatial interaction are imaginable, but often not feasable or possible at all. For example, it is theoretically possible to use gestures (motion patterns) for controlling an application, but due to the inherent hand-screen coupling every motion also results in movement of the screen. Hence, making Wii-like gestures is hardly possible since you would not see the screen anymore: only small gestures are possible, and even these can be cumbersome. Using miniature spatial controllers with a handheld is hard, since place on or next the device hardly permits this. As such, most complex interactions are performed using buttons or pen, using traditional desktop methods.

References

Marsden, Tip. Navigation control for mobile virtual environments. MobileHCI'05. 2005.

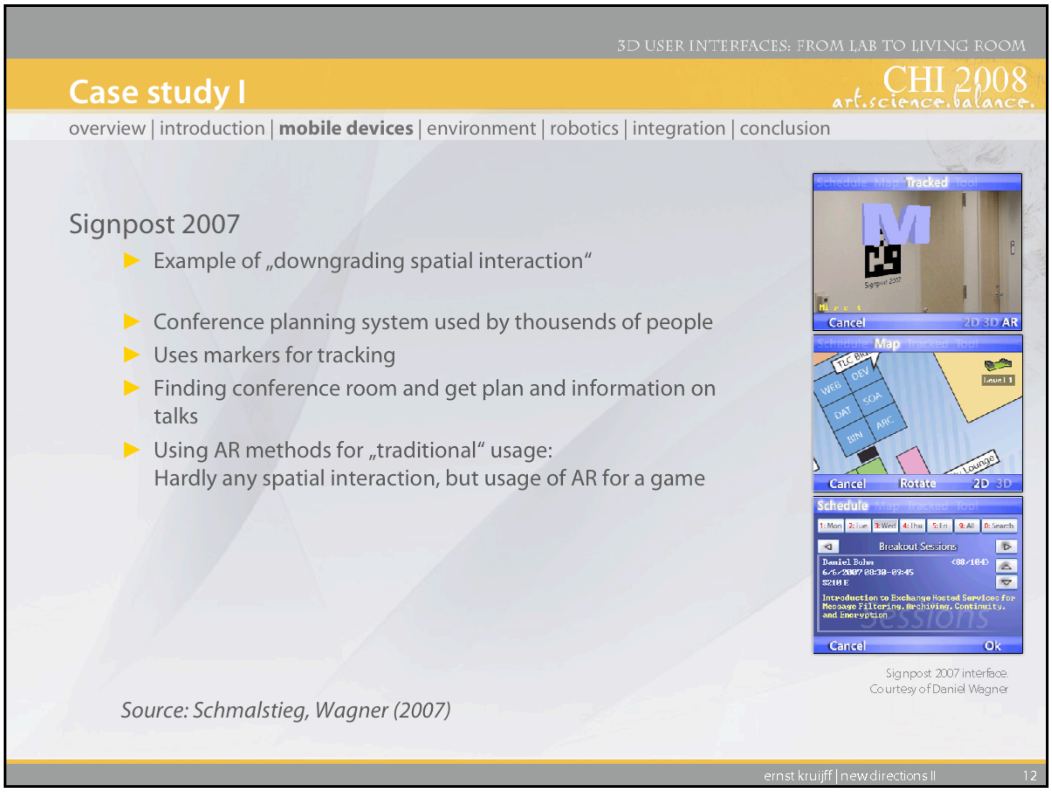
Rohs, M., Zweifel, P.A.: Conceptual Framework for Camera Phone-based Interaction Techniques. In: Gellersen, H.-W., Want, R., Schmidt, A. (eds.) PERVASIVE 2005. LNCS, vol. 3468, pp. 171–189. Springer, Heidelberg. 2005.



A wide range of application examples for MSI can be found. To start with, classical AR applications from the field of engineering are highly applicative to mobile device platforms, as long as the rendering requirements stay within limits. Hence, applications such as on-site checking of piping or electricity lines, or remote maintenance applications can be potentially ported to mobile devices. Next, and already popular, is the usage of barcode encoding of content: widespread in Japan, it already deals with static content, but could easily be extended to deal with pointing to dynamic content that could be downloaded from a remote location using some streaming service. Finally, as pointed out in Roduner 06, mobile device could go into the direction of becoming „universal interaction devices“: though most of the interaction performed with such a device would be non-spatial (programming your VCR, etc.), some examples can be found that could be ported in a useful way to a spatial interaction application. One example is the interactive manual for an appliance.

Reference

Roduner, The Mobile Phone as a Universal Interaction Device: Are There Limits?, MIRW2006, 2006



Signpost2007 is a combination of a conference calendar and a navigation system. The conference calendar can be browsed using various filters, updated by live RSS feeds over Wi-Fi network. All calendar entries are linked to locations: the navigation module can compute the fastest route from the current location (sampled from the last seen marker) to the desired lecture room. The results are displayed on a map that can be freely navigated by panning, rotating and zooming relative to a marker or using phone hotkeys. For large events in venues with multiple levels or buildings, a single map is not enough. Hence, Signpost2007 supports multiple maps linked to a 3D overview, or alternatively an interactive 3D representation of the building showing the global geographic relationship of the current location and the target location. Finally, a built-in Augmented Reality mini-game challenges users with a treasure hunt, that makes use of 3D objects related to each marker in the environment.

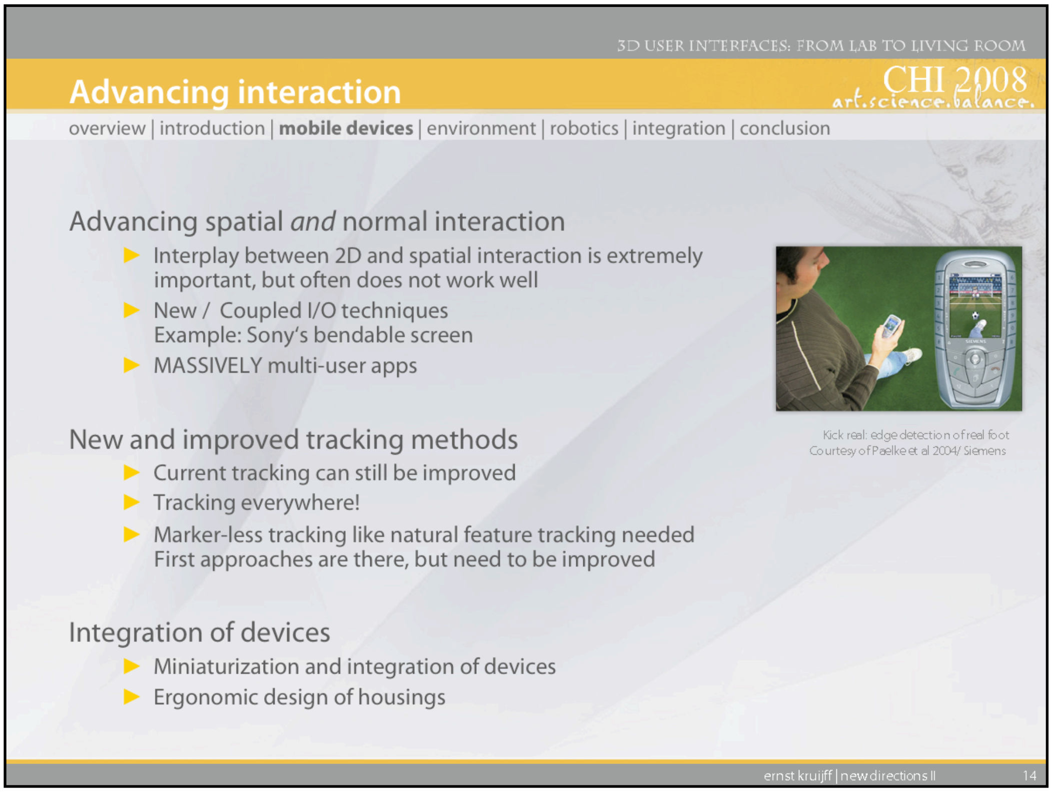
Reference

Schmalstieg, Wagner. Experiences with Handheld Augmented Reality. IEEE ISMAR'07, 2007.

Vesp'R is an ergonomically designed construction to support the interaction with a UMPC using added peripherals. Vesp'R consists of the "BatPack", an enclosure around the UMPC holding peripherals like GPS and orientation sensor, and two joystick-like handles (the "wings") that can be mounted at multiple spots. The devices are made from extremely lightweight ABS plastic (stereolithography) covered by a thin layer of velvety rubber, a hygienic and very soft material to grab. Vesp'R is derived from the Latin word for "bat", a reference the form of the devices, and Ware's "bat" interface studies. The form of the device construction is the outcome of an extensive study on grips, weight balance and micro-electronics to fulfill the needs of outdoor mobile augmented reality. Interface studies have shown that, especially in the two-handed configuration, users can ergonomically hold the Vesp'R over a longer period of time, making use of its range of controllers in an effective way.

Reference

Kruijff, E., Veas, E. Vesp'R - Transforming Handheld Augmented Reality. In proceedings of IEEE ISMAR07, Japan.



In order to advance mobile spatial interaction, several steps need to be undertaken. First, the interplay between 2D and spatial interaction on a mobile device should be more carefully analysed and improved, since it is of utmost importance. As stated throughout the notes, the overlap between mobile interaction and mobile *spatial* interaction is large: many applications will make use of both kinds of techniques. Next, the range of I/O techniques on mobile devices need to be drastically extended. Using the T9/keyboard or micro joysticks is extremely tedious and unnecessary. Hybrid devices merging game devices mobile phones (like the the Nokia NGage is heading for) could have a great future, similar to completely different ways of interaction, such as Sony's concept of a bendable screen (see Jones / Marsden 2006).

Next, tracking techniques can still be improved considerably. „Tracking everywhere" is still an open issue, and not to be achieved by plastering the environment with markers! Also, interest is rising for natural feature tracking, a field which is expected to gain considerable importance over the next years.

Finally, and this is actually an unstoppable process, integration of peripherals in a single device is highly needed. Hereby, hopefully also the quality of embedded systems is improved. For example, some GPS phones exist, but their signal quality is far worse than that delivered by an external device. Also, the ergonomics of mobile devices can still be improved. For example, no matter what mobile phone developers say: I cannot type on a mini keyboard with my big fingers..

Reference

Jones, M. and G. Marsden, *Mobile interaction design*. 2006: John Wiley & Sons Ltd.

Creating interfaces for a mobile device can be a cumbersome process. Over the years, designers have learned from this process and formulated a range of guidelines. A summary of these guidelines can be found here, but it is recommended to read the references and links. The first and probably most defining guideline is to keep the screen size in mind: adapting a desktop interface by just making it smaller doesn‘t work, we don‘t have eagle eyes. Often, one needs to regard minimization of the interface: no screen clutter, less visual menus. Also, data entry should be kept minimal, since on most handhelds it is a hassle. Navigation of any sort (both through menus and through spatial data sets) should be kept simple too. Finally, avoid using too much text on your screen: on smaller screens (outside the UMPC) using graphics works better.

Next, keep the learning curve short for interaction with a device. Complex interaction might be fun to design, but unusable for the user and will therefor often not be accepted (=simply disregarded). Also, spatial interaction is not a „must“ on a mobile device. In many cases, a normal GUI like interface might work much better. For example, why navigate through menus using gestures when you can far more easily do it with a micro joystick? Often, speed is a main factor: users want to do things quick, so don‘t bother the user too much: design for enjoyment, not for stress. This also means that one often needs to deal with split attention: the mobile phone is often used quickly in between other tasks.

Finally, and this is not always possible, try to use „good“ hardware or think about extending a device. For example, using an external camera with a UMPC can greatly improve tracking performance. But, be warned not to destroy the ergonomics doing that – just attaching some periphersals somewhere on the device is not a good idea!

References: see next page

Guidelines II

References

- Gong, Tarasewich. Guidelines for Handheld Device Interface Design. Proceedings of the DSI 2004 Annual Meeting.
- Karampelas, P., Akoumianakis, D., & Stephanidis, C. (2003). User interface design for PDAs: Lessons and experience with the WARD-IN-HAND prototype. In N. Carbonell, & C. Stephanidis (Eds.), *Universal Access: Theoretical Perspectives, Practise and Experience - Proceedings of the 7th ERCIM Workshop "User Interfaces for All"*, Paris (Chantilly), France, 24-25 October (pp. 474 - 485).

Links (from Wagner et al 2007)

- Do's and Don'ts of PocketPC design / http://www.pocketpcmag.com/_archives/Nov04/Commandements.aspx
- Usability special interest group – handheld usability / http://www.stcsig.org/usability/topics/handheld.html
- Usable Mobile website / http://www.smartgroups.com/groups/usablemobile
- Mobile Coders Website / http://www.mobilecoders.com/Articles/mc-01.asp
- Univ of Waikato Handheld Group / http://www.cs.waikato.ac.nz/hci/pdas.html
- Mobile Interaction Website / http://www.cs.waikato.ac.nz/~mattj/mwshop.html

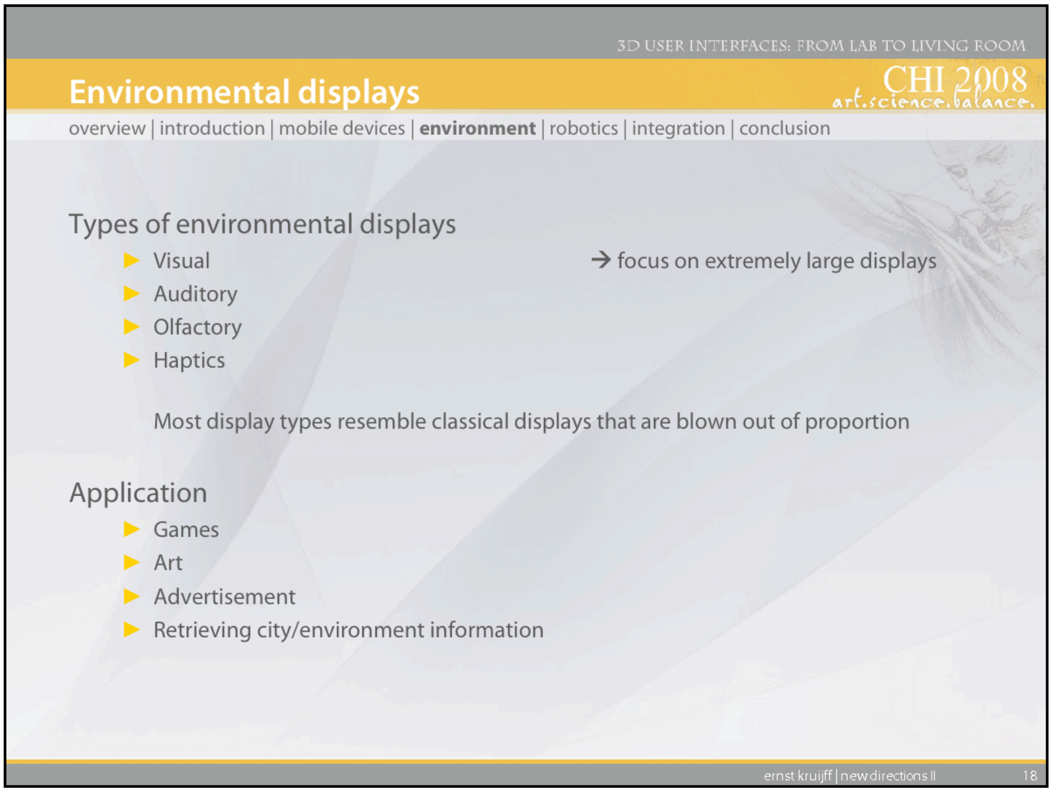
Recommended reading

Books on handheld interaction

- Jones & Marsden, Mobile Interaction Design
- Zwick et al, Designing for Small Screens
- Weiss, Handheld Usability

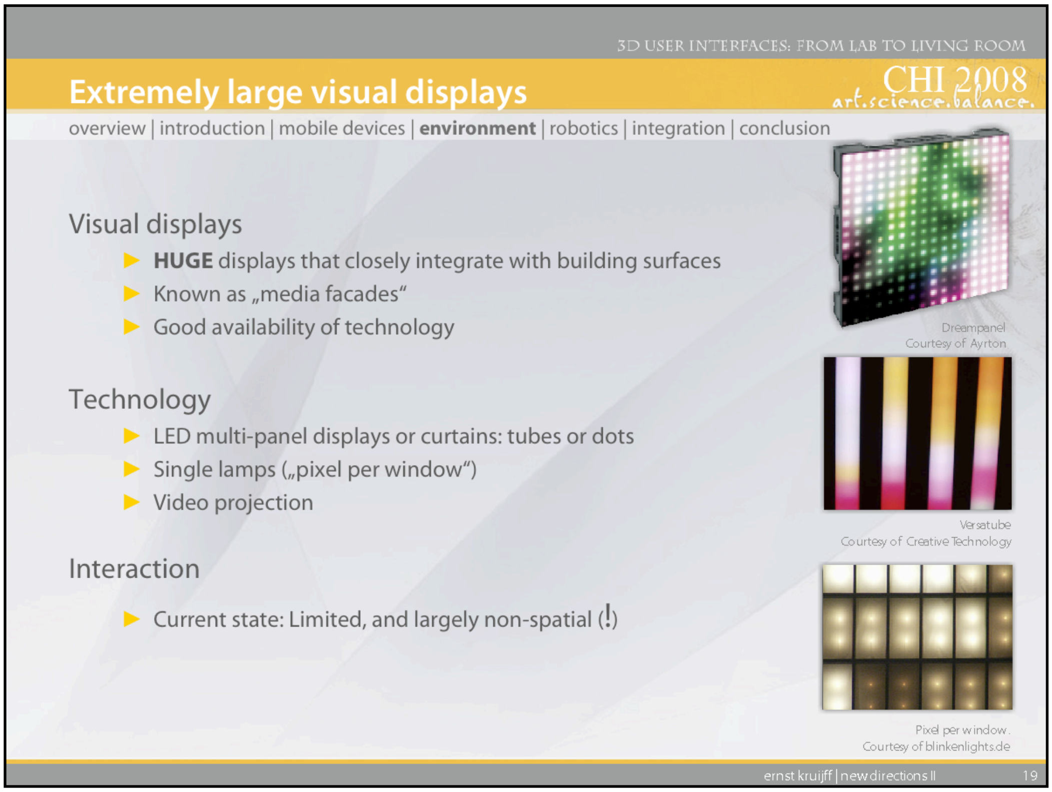
Course

- Wagner, Barakonyi, Billinghurst, MacIntyre
  Designing and Developing Handheld Augmented Reality Systems.
  IEEE ISMAR 2007.
- CHI courses

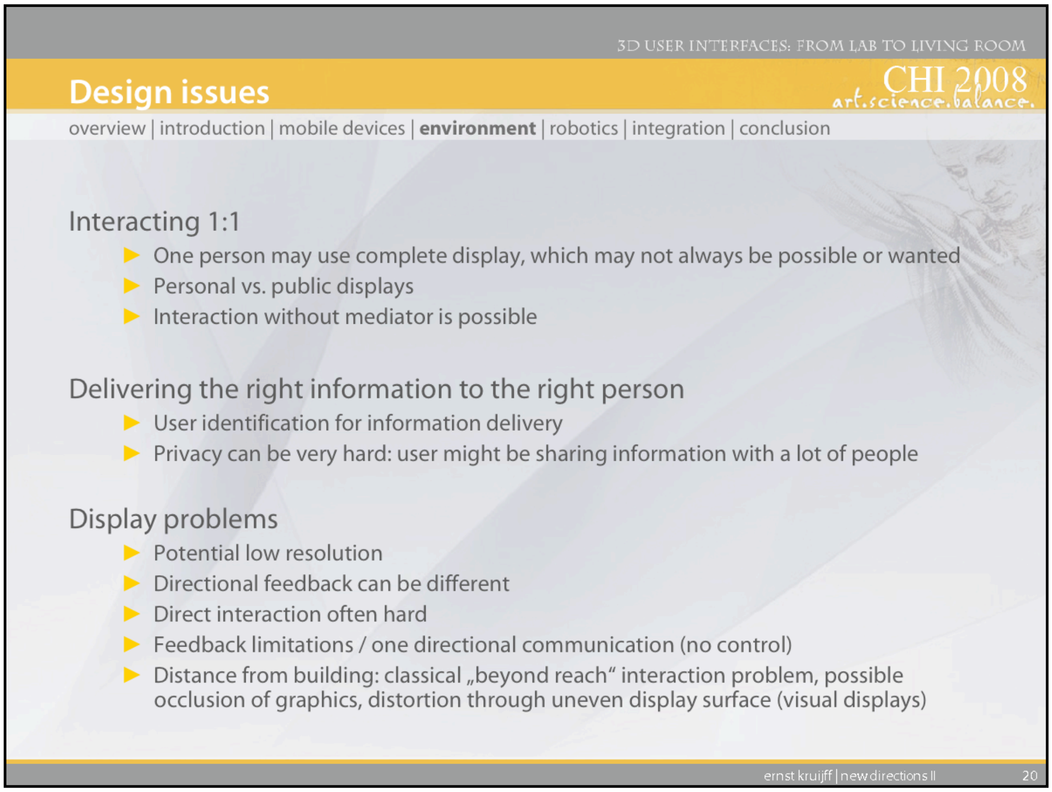


This part of the talk deals with environmental displays. These kinds of displays are embedded in the world around us – the main focus is on large scale display systems, not small ones: we don‘t focus on normal terminal-like systems. One might see this part of the talk as the „next stage of billboard advertisement“ – all display types presented here resembles classical display systems in one way, but are blown out of proportion. Whereas we mostly talk about extremely large *visual* displays, we also take a quick look at possible auditory, olfactory and haptic displays.

Most large scale display surfaces are currently used for artistic or game purposes, but potentially, their field of application is wider.



Extremely large visual displays are the most widespread kind of environmental display. Over the years, the have come to be known under the name of „media facade“, and can be built up well through availability of a good amount of systems. Most systems (like from Ayrton, Barco or Creative Technology) make use of LED technology, put into some panel structure in which multiple panels can be coupled. Some approaches have also experimented with single lamps behind windows (see case study), likely the cheapest way to make a media facade. Both LED panels and lamps can be directly integrated in a building, or is even used as defining architectural element. A final possibility is projection: Especially due to the availability of high quality (resolution and brightness) projectors, systems also make use of external projection.



Using huge displays brings upon several design issues that need to be regarded. Especially with visual display systems, it may often happen that only one person can interact, which is often a restriction caused by badly available multi-user interaction techniques. The next point is the matching of information to the right user, and the privacy / social issues involved in that, which might be a serious problem that may require techniques like filtering of content.

Even when the content gets to the right person, the display systems have inherent limitations that make interaction, especially spatial interaction, a hard topic. Displays can have a very low resolution, and specifically with visual displays, there wíll be a classical „beyond reach" problem (users cannot directly interact with objects), next to possible occlusion of graphics (by other buildings or other city infrastructure like billboards). Furthermore, massive distortion of the displays may occur, due to the size or form of the display surface.

Finally, direction of feedback is a hard topic – feedback possibilities are limited with all kinds of displays. Lateron, we will see how some of the directional issues are handled.

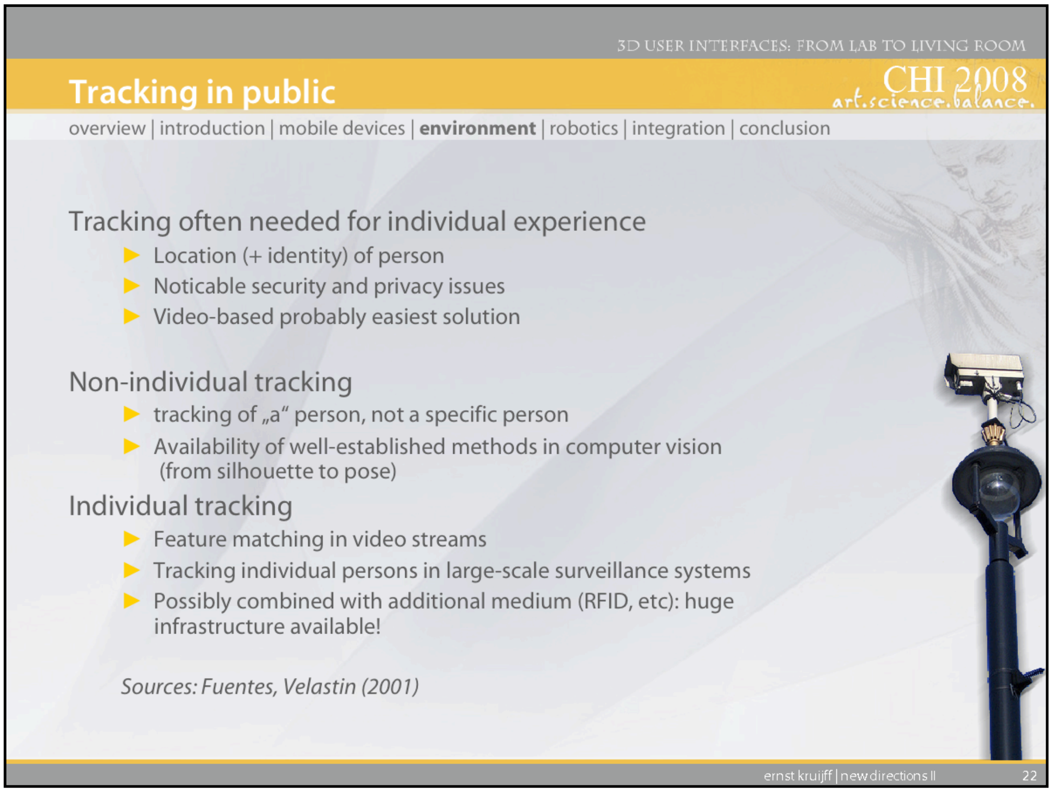
Display examples

Shibuja, Tokyo, Light advertisement.
Courtesy of Kruijff

Uniqua Tower, Vienna, using 2800 MIPIX (LED) pannels.
Courtesy of Barco

Easyweb projection van with 10K lumen projectors
Courtesy of easyweb / Neoproj.

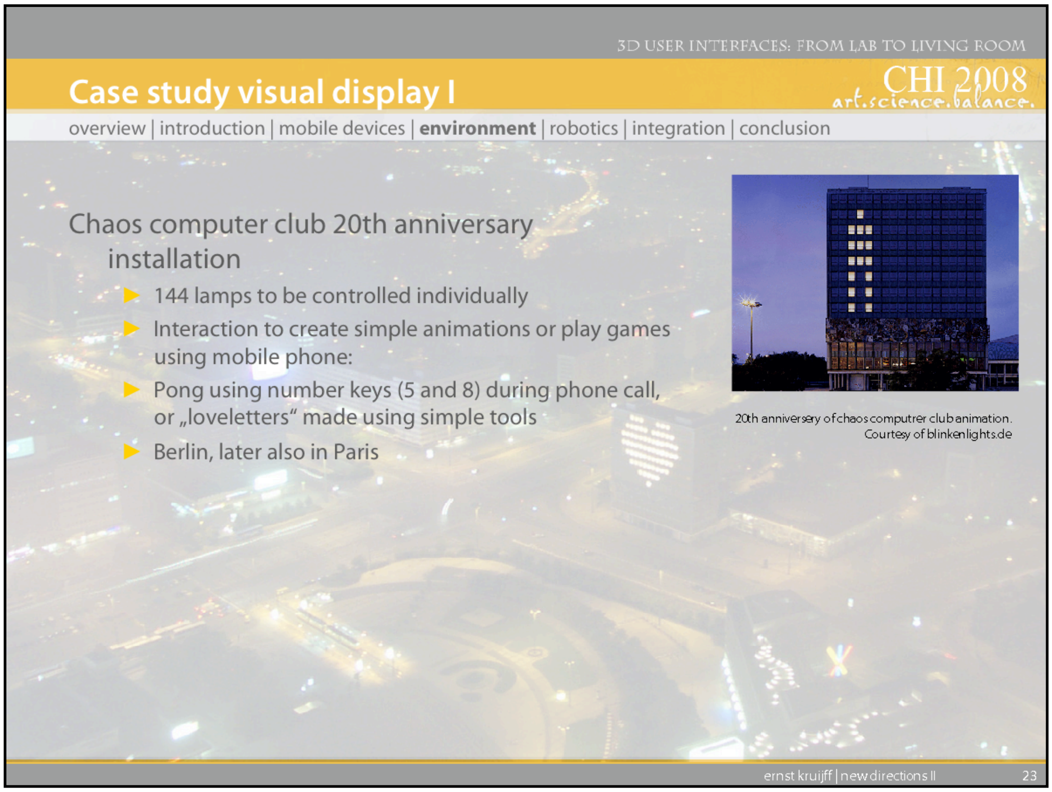
Maxalot projection, using 20K Lumen projectors.
Courtesy of Maxalot



In order to support an „individual experience“ in an environmental display system, it will be needed to somehow locate and possibly identify the user. Naturally, this leads to considerable privacy, social and security issues. The tracking will most likely be based upon camera-based methods, simply because it is a rather straightforward method with a widely available infrastructure. Identifying a person can be eased by using IDs like RFID. However, tracking of individual persons is already possible in setups like coupled surveillance systems, by using feature matching techniques on video streams. Easing security issues, one can also track persons without identifying them: silhouette tracking is a well-established method in the field of computer vision and could easily be used for this purpose (and has, actually, already been since installatons like Videoplace at the start of the eighties). It is currently unlikely that any other tracking technique can be as flexible as computer vision based tracking for localization persons in large spaces.

Reference

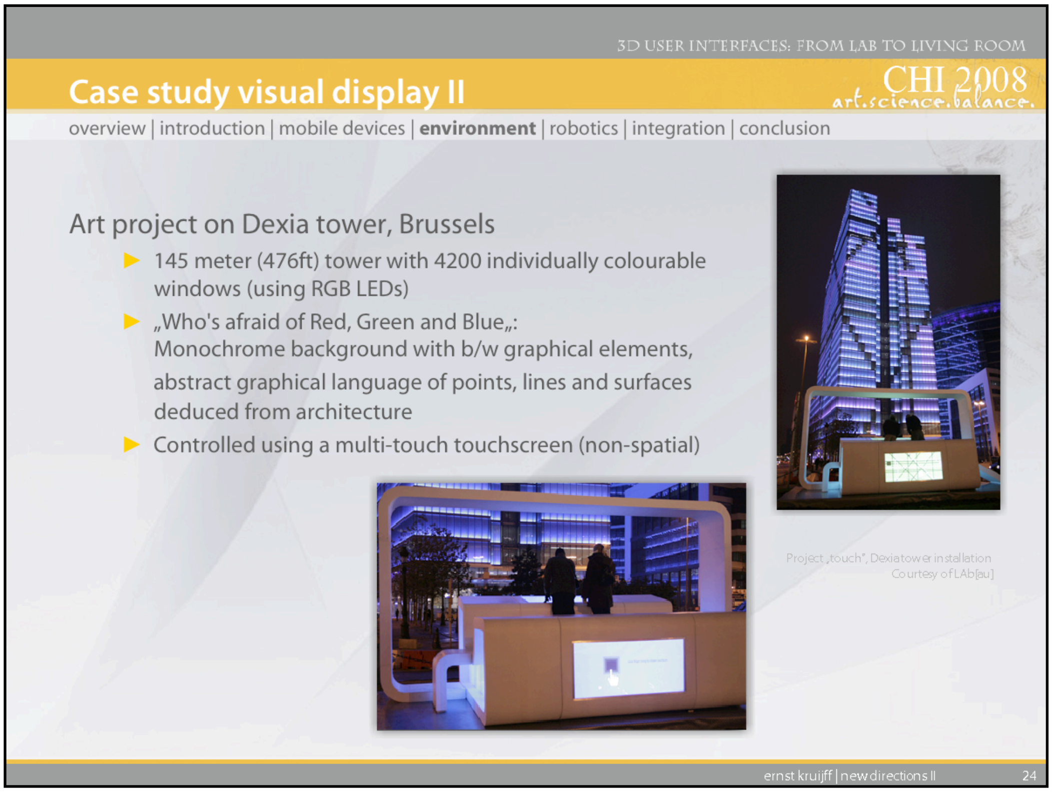
Fuentes, Velastin. People tracking in surveillance applications. 2nd IEEE International Workshop on Performance Evaluation on Tracking and Surveillance, PETS 2001, Kauai (Hawaii-USA), (2001)



One of the best known and almost a classic by now is the installation made by the Chaos Computer club for their 20th Anniversary. In 2001/2002, the famous "Haus des Lehrers" (house of the teacher) office building at Berlin Alexanderplatz was enhanced to become world's biggest interactive computer display: Blinkenlights. The upper eight floors of the building were transformed in to a huge display by arranging 144 lamps behind the building's front windows. A computer controlled each of the lamps independently to produce a monochrome matrix of 18 times 8 pixels. During the night, a constantly growing number of animations could be seen. But there was an interactive component as well: one could play the old arcade classic Pong on the building using a mobile phone or place lovelettes on the screen as well.

Link

http://www.blinkenlights.de

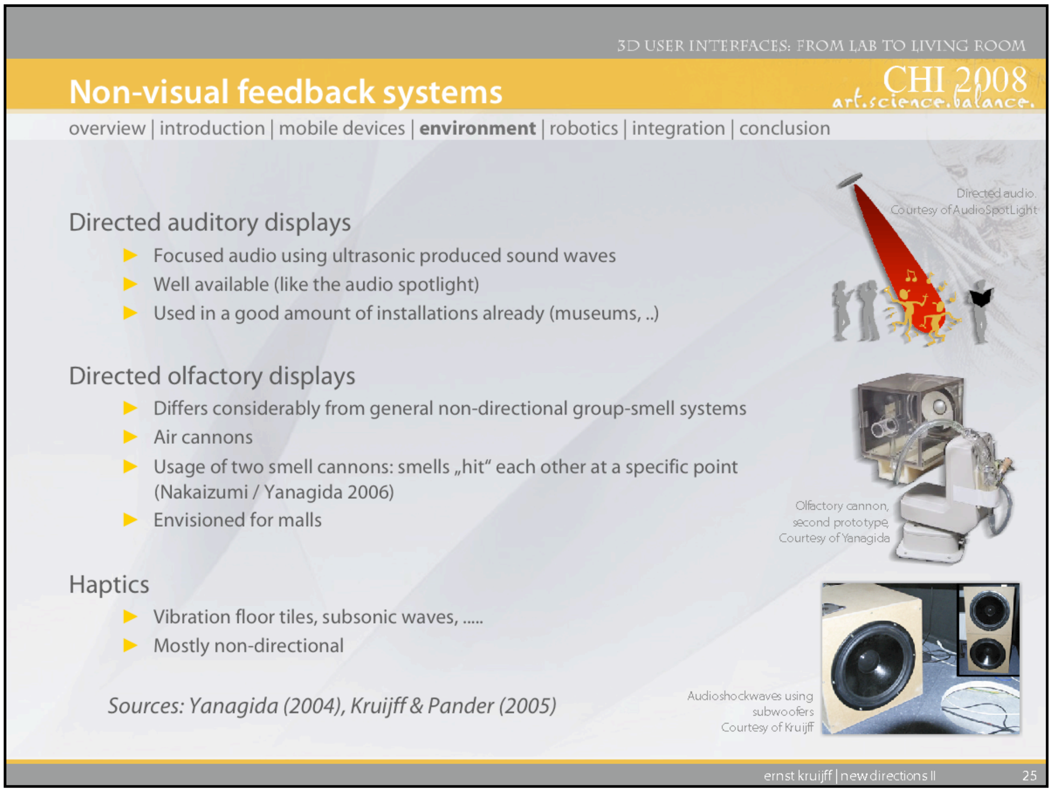


Since 2007, an art project was presented using the Dexia tower in Brussels as huge artwork. The 145 meter high tower holds the massive amount of 4200 individually controllable RGB LEDs, that are used to display the artwork „Who‘s afraid of Red, Green and Blue“. The work represents an abstract and geometric language based on points, lines and surfaces, expressing the progression of time from sunset to sunrise. The intensity of light and / or density of shapes evolves according to the progression of time, midnight being the shift between the current day and the next. In this manner, the progressive increase / decrease of the tower enlightening inverts the logic of day=light / night=darkness having its cumulating point at midnight and its lowest level in the morning. Finally, the relation of color to weather forecast establishing weather as another parameter of light (Lab [au] website). The pattern shown on the tower is controlled by a straightforward application on a touch-screen located close to the tower.

Copyright / reference:

Artists: LAb[au]
Architects: Philippe Samyn & Partners, M & J.M. Jaspers - J. Eyers & Partzners
Lightning engineer: Barbara Hediger

http://www.lab-au.com



Next to visual displays, there are some other techniques that can potentially cover large scale environments. On this slide, we present some of these techniques. First, several audio systems exist that can be used well for delivering personalized audio experiences in public spaces. Systems exist that make use of ultrasound to create beams of audio that can only be heared by the persons standing within range of the beam, hence dealing well with the previously mentioned „directional feedback“ issue.

Another kind of feedback towards a user that can theoretically also be created „on the spot“ is using air cannons to generate smell sensations (Yanagida et al 2004). By coupling two air cannons, one can also generate approximate spatially located smell events: the smell „balls“ hit each other on a predifined trajectory. This kind of system greatly differs from general smell systems that provide a smell sensation to a larger space, for example using airconditioning systems.

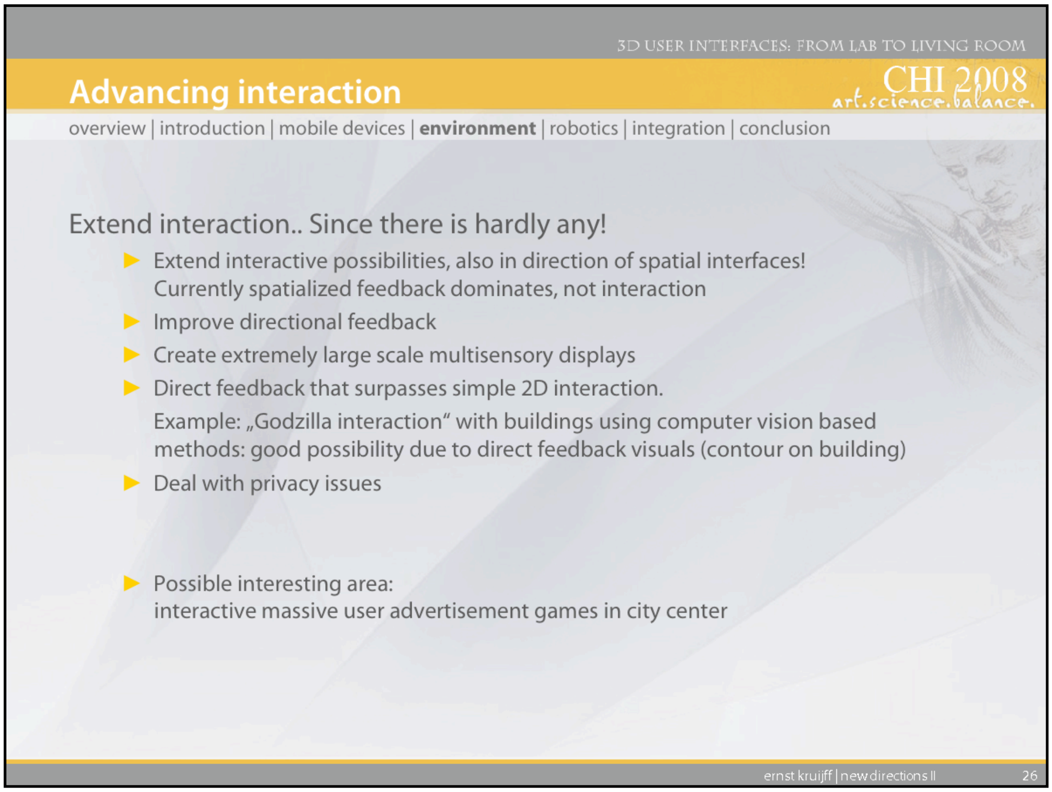
Finally, the generation of haptics is extremely hard for larger groups of people, or in larger spaces. Some approaches exist that make use of vibration elements in floortiles to generate haptic-like feelings. Another approach is taken by generating low frequency shock waves. As can be noticed in rock concerts, low frequencies generate vibrations in empty volumes in the body (especially lungs) that feel like a haptic sensation. As such, using large subwoofers feedback can be provided to large groups of people. Informal tests did not show nausea with users (Kruijff & Pander 2005).

References

Audiospotlight: http://www.holosonics.com/

Kruijff, E., Pander, A. Experiences of using shockwaves for haptic sensations. In proceedings of 3D user interface workshop, IEEE Virtual Reality Conference, Bonn, Germany, 2005.

Nakaizumi, Yanagida, Noma, Hosaka. SpotScents: A Novel Method of Natural Scent Delivery Using Multiple Scent Projectors. Proceedings of IEEE Virtual Reality 2006, Alexandria, Virginia, USA, pp. 207-212, March 2006.



Whereas technology-wise, large scale environmental displays could be created, one hardly sees any spatial interaction. Appropriate spatial interaction techniques need to be found that can cope with technical characteristics of the displays. In addition, directional feedback techniques for large spaces can still be advanced to a large extend. One idea is to blow up the proportions of Sony Eye-Toy like interaction to create a kind of „Godzilla interaction" in which one sees his/her owns solhouette on a building to interact with objects – certainly, this kind of interaction will look very impressive. Also, this kind of interaction can easily be extended to multi-person interaction.

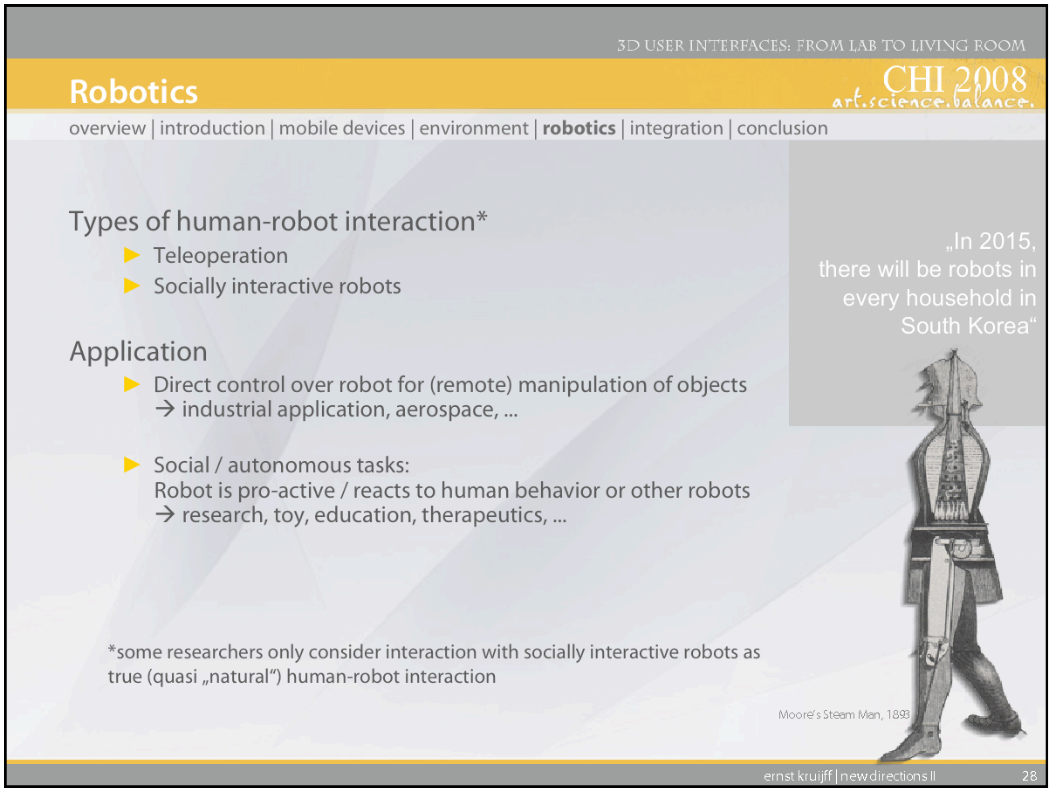
Recommended reading

Literature is extremely sparse

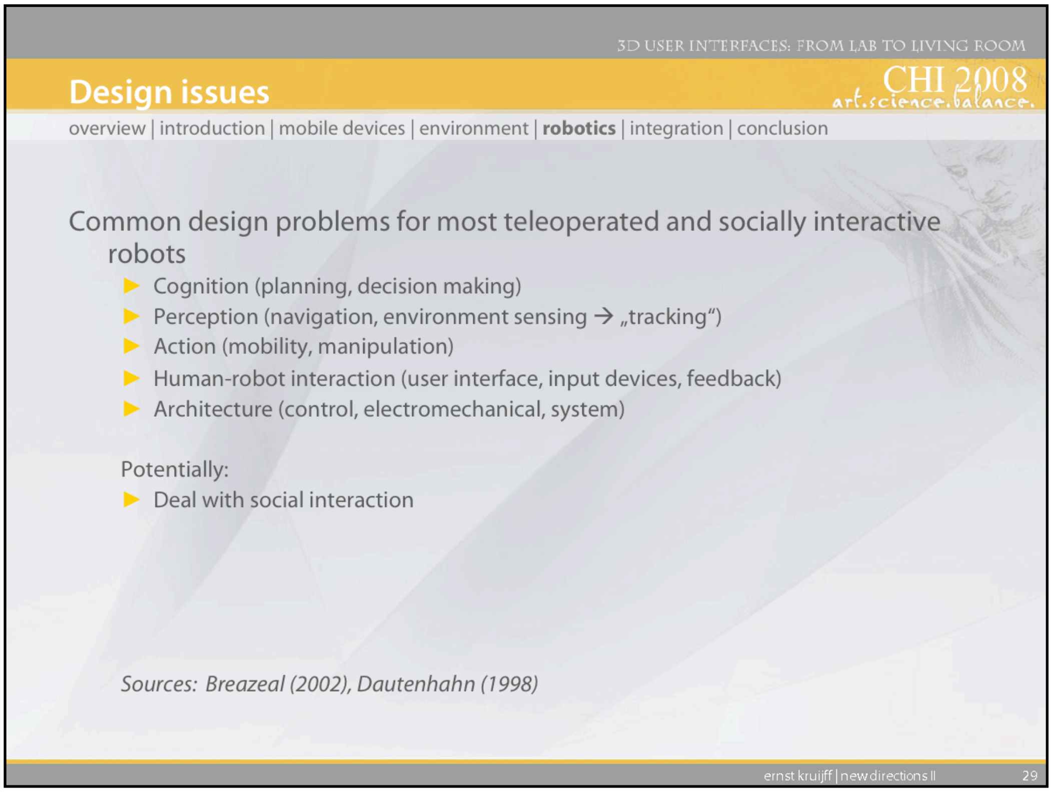
- This must be an open research area....

Book

- Ag4 media facades, daab, 2006 / www.mediafacade.com



The third part of this talk takes a look at different types of human-robot interaction. Whereas teleoperation-type interfaces are predominantly used in industrial applications, the rise of more simple house robots, up to the ones that are even socially interactive is likely change the popularity of robotics considerably. In many countries, the availability and usage of robots is believed to blossom in the following decade. Robots can be controlled directly, but autonomity of robots is rising: socially interactive robots exist that are pro-active to human actions / behavior to perform tasks alone or with other robots. Here, the fields of application go beyond industrial applications: robots are already used as toy, for education and as therapeutic aid.



Most teleoperation and socially interactive robots have some common design problems that need to be tackled. They require some cognitive component for decision making (especially the socially interactive robots), and some perceptive mechanisms for navigation and environment sensing. In addition, most robots need some kind of action mechanisms with which they can interact with their environment, may it be a method to move around or specific manipulation aids like a mechanical hand. The ways of human-robot interaction can differ, as we will see in this talk, ranging from almost complete control by a human up to largely autonomous behavior. Finally, there will be the overall system architecture that binds all components.

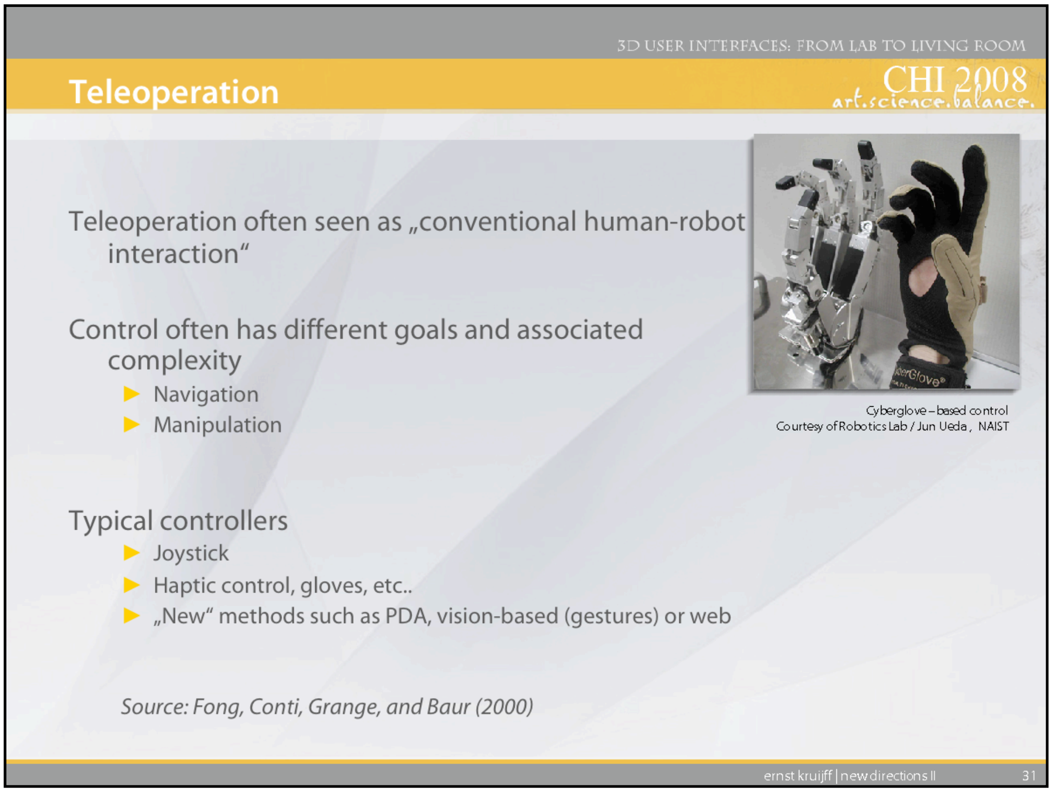
For the socially interactive robots, there is an additional point: deal with social interaction itself. This often requires some kind of „deep model“ for human interaction to encourage communication between users and robots.

References

C. Breazeal, Designing Sociable Robots, MIT Press, Cambridge, MA, 2002.

K. Dautenhahn, The art of designing socially intelligent agents—science, fiction, and the human in the loop, Applied Artificial Intelligence Journal 12 (7–8) (1998) 573–617.


3D USER INTERFACES: FROM LAB TO LIVING ROOM
Tracking
CHI 2008
art.science.balance.
Robots mostly make use of computer vision based methods, possibly coupled to GPS (SIR), teleoperation mostly mechanical
► Computer vision methods for robot perception well established
ernst kruijff | new directions II
30

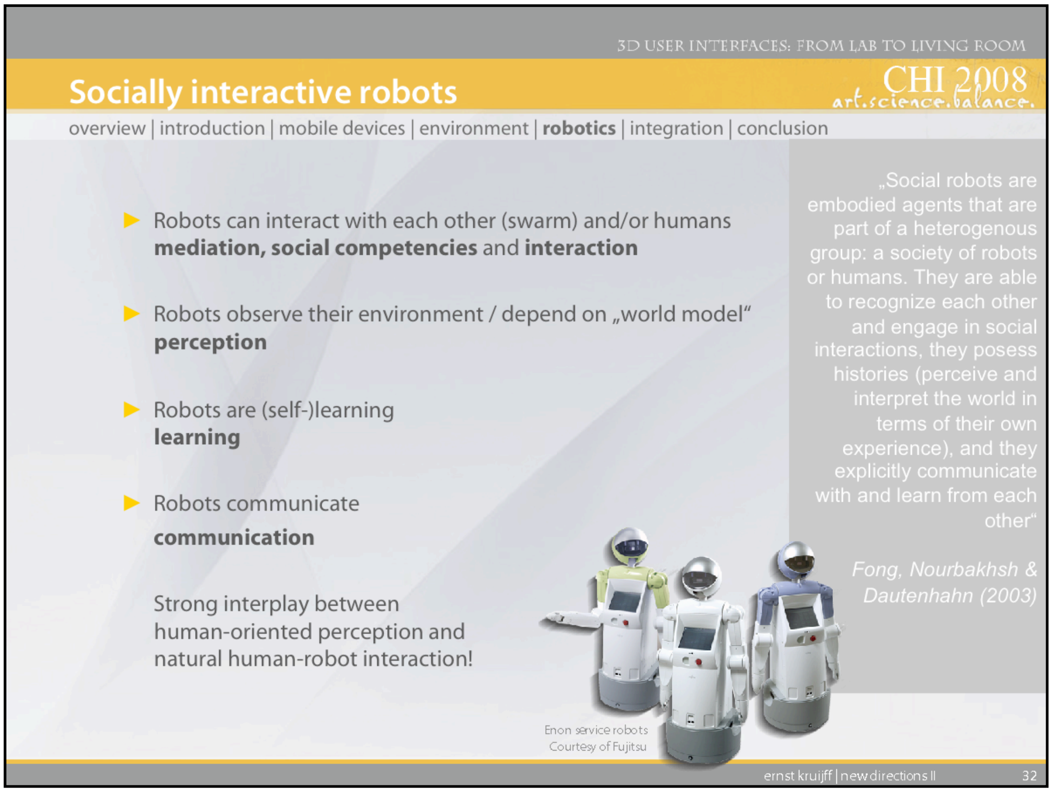


Teleoperation is generally seen as the „traditional“ field of robotics, providing the more conventional ways of control / interaction. Though this vision migth be true with regards to the maturity of the field (interfaces have been around for a long time), they still show a considerable amount of innovation. On the other hand, they completely lack the „intelligence“ and behaviouristic characteristics of socially interactive robots, that are by now receiving far more attention.

As the word teleoperation implies, controlling basically means mimicking human movement. This may happen by a simple device like a joystick or some nobs, up to more advanced haptic devices or gloves. Furthermore, some new methods using a PDA or computer vision have appeared. These methods have not found much usage yet, even though they can (and have) also be used for socially interactive robots.

Reference

Terrence Fong, Francois Conti, Sebastien Grange, Charles Baur**.** Novel interfaces for remote driving: gesture, haptic and PDA. SPIE Telemanipulator and Telepresence Technologies VII, 2000



Socially interactive robots (SIRs) can act alone, in groups, or with humans to solve some kind of problem. Often, they serve as mediator between a user and the environment to perform some task, but there are also robots that can be completely autonomous. For example, think about a cleaning robot that just cleans every couple of days independant of the user, unless specifically requested to clean a specific area after some kind of accident. In any case, the robot needs to be socially competent. Social competence can occur at different levels, basically based on the level of „intelligence“ the robot inhabits, which often relates to the task domain it operates in.

Robots observe their environment continuously using a variety of sensors, and highly depend on the world model that defines the context of action. In relation to its overall cognitive architetcure, this world model may change under the process – robots tend to learn of the tasks / actions they perform. Observation takes place at different levels: observing the environment itself (localization, object avoidance, ...), sense and interpret the same phenomena a user observes, and finally, observe the user him/herself to built up a useful and natural kind of communication.

Reference

Fong, Nourbakhsh, Dautenhahn. A Survey of Socially Interactive Robots. Robotics and Autonomous Systems, 42. 2003.

Behavior is one of the key issues to make a SIR. It is of importance to built up a social relationship between the user and the robot, requiring a cognitive architecture to create a natural model of communication. A cognitive architecture needs to represent and execute motor skills to operate in an environment. In order to create a natural behavior, the robot at least requires a good level of situational awareness to act appropriately. It needs to recognize situations and events based on familiar patterns – it needs to able to categorize (map) objects, situations and events to concepts. As such, the cognitive architecture needs to consist of a representation and systematic organization of patterns, and measures to match situations to these concepts. During interaction with its environment and users, the robot will learn and adapt or create new patterns.

All together, the cognitive architecture largely resembles human problem solving and action planning mechanisms. Thus, knowledge from these domains can be used to model the cognitive architecture in a valid way.

Reference

Kruijff, GJ. How to make talking robots. Tutorial at RO-MAN’07.

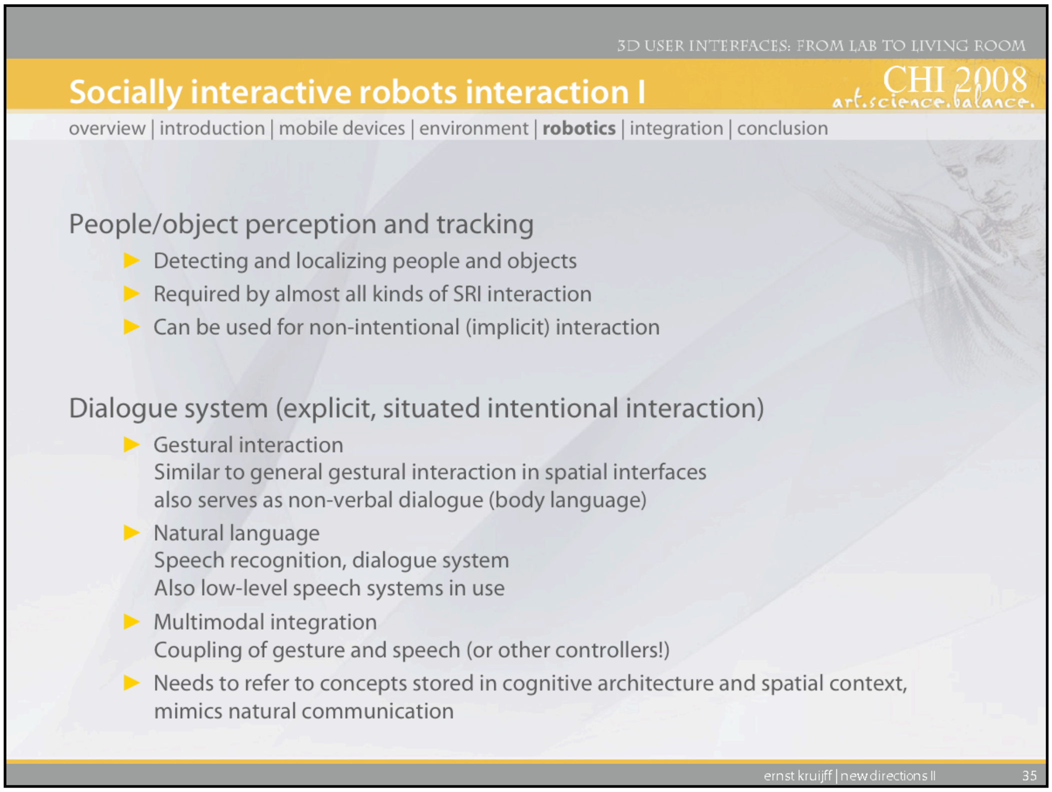
Emotion can seriously advance the robot behavioristic, communicative and interactive capabilities. However, emotion is a very difficult topic and mostly handled at top-level by sorting emotions in just a few categories. Emotion can be a two-way street: the robot should be able to recognize human emotion and possibly express emotion itself. Recognizing mostly happens on the base of face recognition by detecting a face, extracting its features and classify the expression. In addition, voice patterns or gestures (body language) are regularly used to detect emotion. In order to communicate expression, the robot may be able to create different facial expressions, or provide some speech or gestural feedback. Emotion highly depends on the appearance of the robot. Here, different directions are taken, including biologically oriented robots, robots that look like animals, or more abstract versions that may resemble a caricature / cartoon figure. Especially with the biologically oriented direction, problems exists with the acceptance of a robot: a range of visual appearances exist that human just interpret as „scary“, something which Mori found out in the eighties. Quite a number of experiments are being done to find out how humans truly react to a robots experience, especially since robots can look extremely similar to human beings themselves. As can be seen on the picture in the slide, it may take a closer look to identify that the left figure in the picture is not a human being.

References

C. DiSalvo, et al., All robots are not equal: The design and perception of humanoid robot heads, in: Proceedings of the Conference on Designing Interactive Systems, 2002.

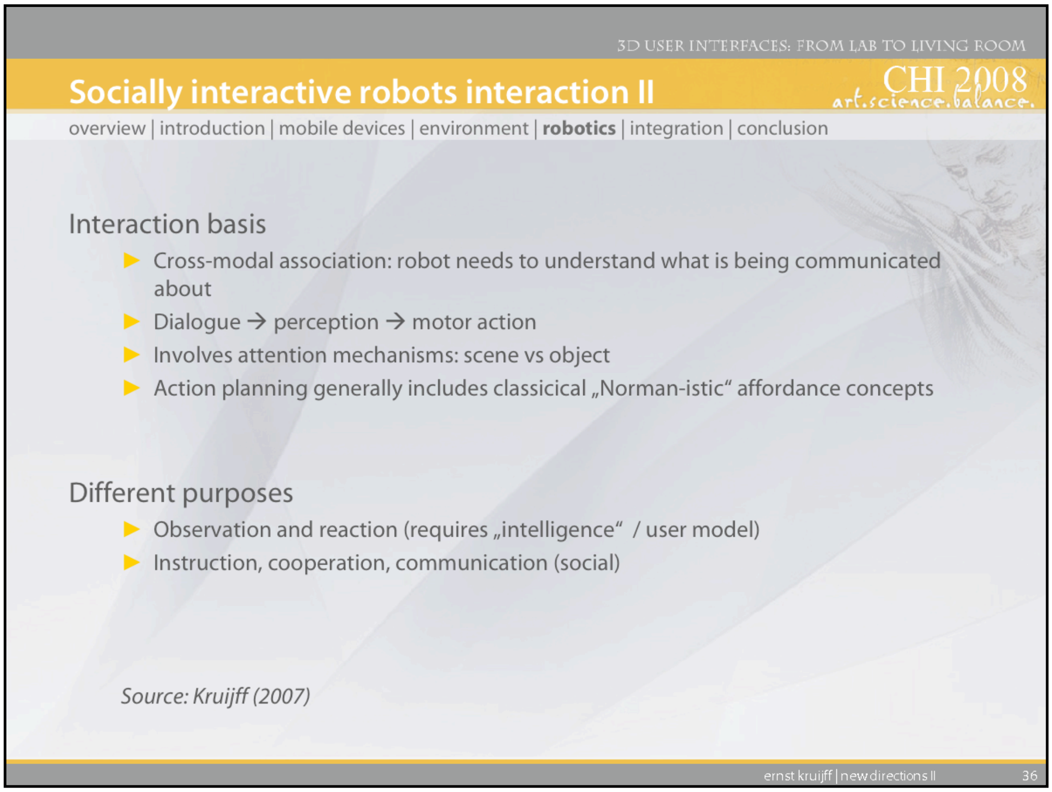
K. Dautenhahn, C. Nehaniv, Living with socially intelligent agents: A cognitive technology view, in: K. Dautenhahn (Ed.), Human Cognition and Social Agent Technology, Benjamin, New York, 2000.

P. Persson, et al., Understanding socially intelligent agents— A multilayered phenomenon, IEEE Transactions on SMC 31 (5) (2001).



In order to allocate interaction between a user and a robot, the first step is to perceive and track a person, and possibly some related / associated objects. People detection is a well-dealt field in computer vision, and can be used for both direct interaction and non-intentional (explicit) communication between a robot and a human, once features of a human are detected that help categorization of body parameters.

Direct communication and interaction with a SIR mostly takes place via some form of dialogue system. Here, interaction is explicit and depending on the context the user and robot are in. Dialogue systems are mostly based on traditional language dialogue and / or gesture systems to recognize content, which needs to be placed within the context of action. As such, the dialogue needs to refer to both the concepts stored in the cognitive architecture, and an identified spatial environment.



The previously mentioned notion of context mostly coincides with cross-modal association, in which the robot needs to combine different sources to understand what is communicated about. Hence, dialogue leads to perception, after which a motor action is planned: this kind of process is highly similar to any perception-action cycle humans go through, and involves specific attention mechanisms (like scene vs object) that are similar to what we see in human actions too.

When the robot plans an action, traditional ideas like the in spatial interaction well known „Norman-istic“ concepts of affordance appear: the robot needs to understand what kind of actions it may be able to perform with an object. Whereas this may be any easy task for a human being, it can be extremely hard for a robot.

Robot actions can have different purposes. The robot may observe and react to an environment that may inhabit users, which requires some kind of intelligence, and possible a user model. A user model contains attributes that describe a user, or a group of users and can be both static or dynamic (via learning). Direct interaction with users or other robots can aim at plain communication (social action), but of course can also trigger instructional tasks that can be based on a plain command structure („do that“) or may include cooperative aspects.

Reference

Norman, D. The Design of Everyday Things. 1988.

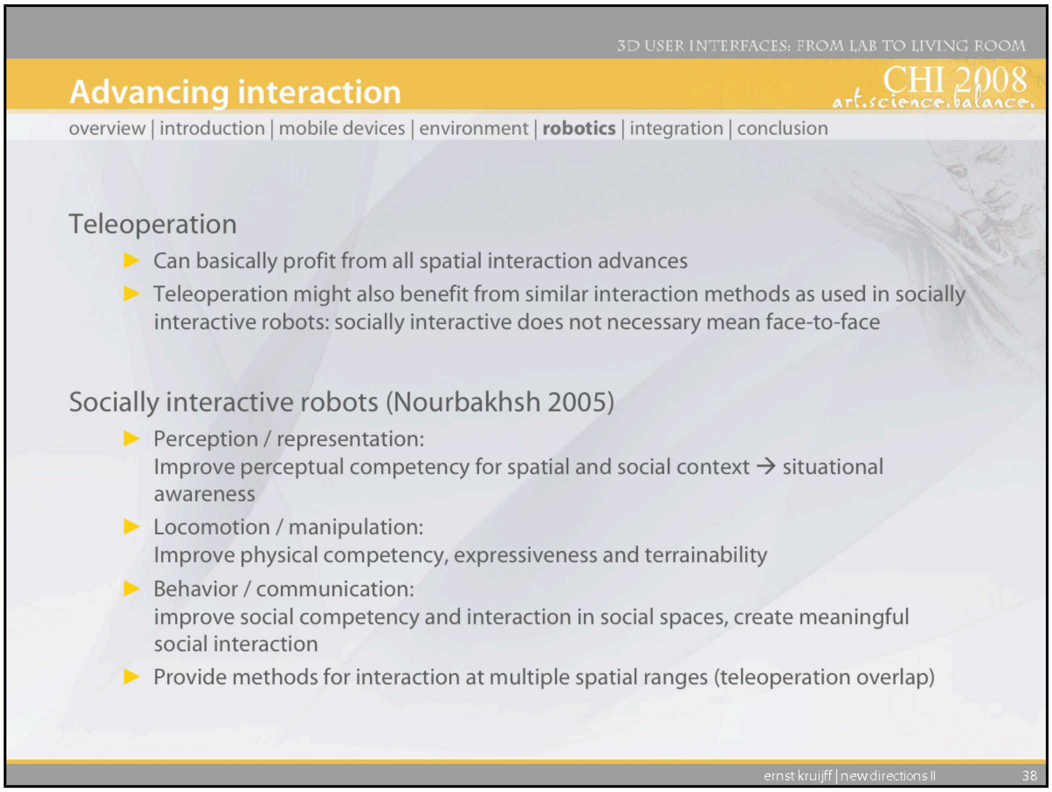
Kruijff, GJ. How to make talking robots. Tutorial at RO-MAN’07.

RoboCup@Home is a new league inside the Robocup competitions that focuses on real-world applications and human-machine interaction with autonomous robots. The aim is to foster the development of useful robotic applications that can assist humans in everyday life. The league is especially interesting since it brings together a multitude of tasks that require careful development of the different components talked about before. The robots need to perform completely autonomous (no cabling or direct control), but may be instructed by the user via dialogue. The tasks include different actions such as following someone, or finding a specific (learned) object in the environment. The environment itself currently consists of a modeled living room, but extensions are foreseen, like a search task in a supermarkt.

Link

http://www.ai.rug.nl/robocupathome/

http://www.flea.at



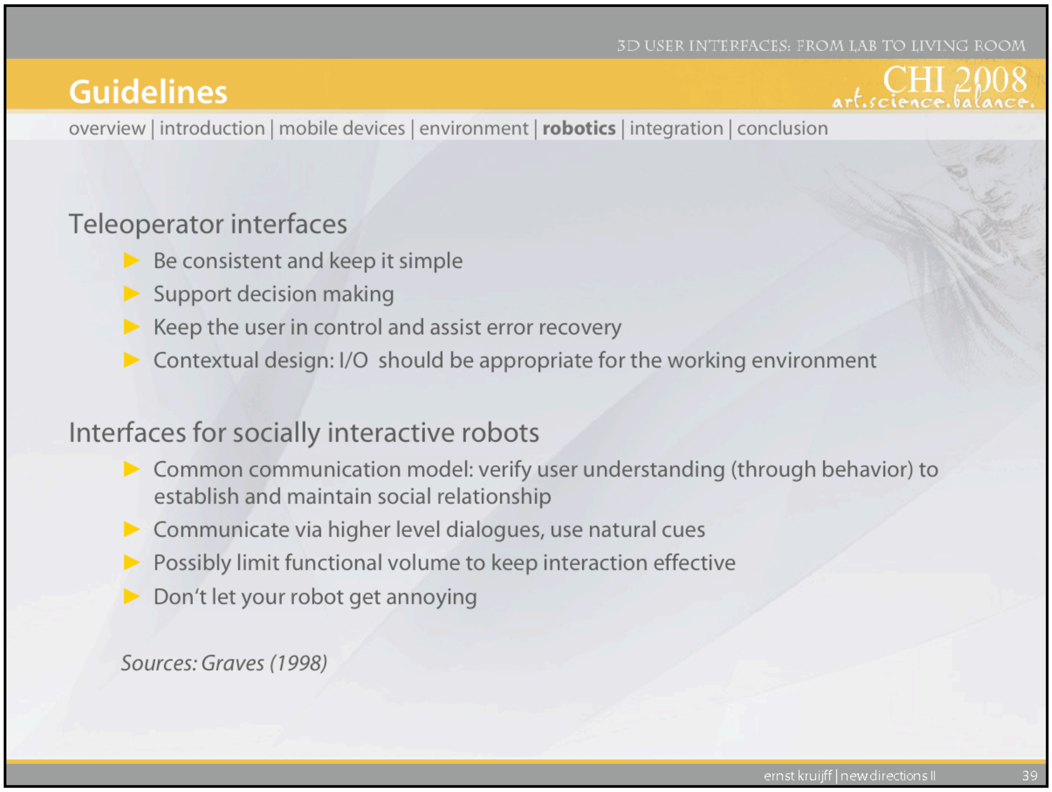
Human-robot interaction is an active field of research that can still advance in multiple directions. First, teleoperation can profit from most developments in the general field of 3D user interfaces, due to its strong overlap with standard manipulation tasks in 3D environments. In addition, teleoperation may learn or even converge quite a bit with SIRs. Socially interactive does not necessary mean face-to-face: remote operation of a robot may also be based on dialogue.

Next, as identified by Nourbakhsh, SIRS still have major challenges to go through. The robots perceptual, physical and social competencies can still be improved a lot. Finally, feeding back to the previous comment on overlap with teleoperation, SIRs require methods that can deal with multiple spatial ranges. Interaction may not necessarily be face-to-face, but like within space exploration happen on three levels: close, within view, over the horizon. For such scenarios, the overlap with teleoperation is considerable.

References

Fong, Nourbakhsh. Interaction challenges in human-robot space exploration. ACM Interactions, Volume 12 , Issue 2, 2005.

Nourbakhsh. The Wicked Problem and Interaction Evolution. CMU course notes, 2005.



When creating teleoperator or SIR interfaces, several guidelines should be taken into regard.

Teleoperator interfaces ultimately need to allow for simple and direct interaction with objects, and support decision making while doing so. At all times, the user should be in control of what happens – hereby, it is important that methods exist that aid in quick error recovery. Finally, I/O should always be choosen by thinking about the final working environment: of course, space exploration has different requirements that using a robot in a factory.

Regarding SIRs, it is always important to create and keep a common communication model between the user and the robot, and to regularly check if the user and robot are still understanding each other. Without, a social relationship between user and robot is extremely hard to establish. In addition, it may be required to communicate via higher level dialogues and to use natural cues (like human expressions mimicked by a robot) to confirm communication understanding. With many robots, it can be required to limit the functional possibilities to keep the interface effective. Finally, avoid that the robot becomes a nuisance, for example by deadlocks in dialogues or following users at too close range: robots should be „non intrusive“.

Reference

Graves, A. 1998. User interface issues in teleoperation. Working Paper 3, Centre for Computational Intelligence, De Montfort University, Leicester, United Kingdom.

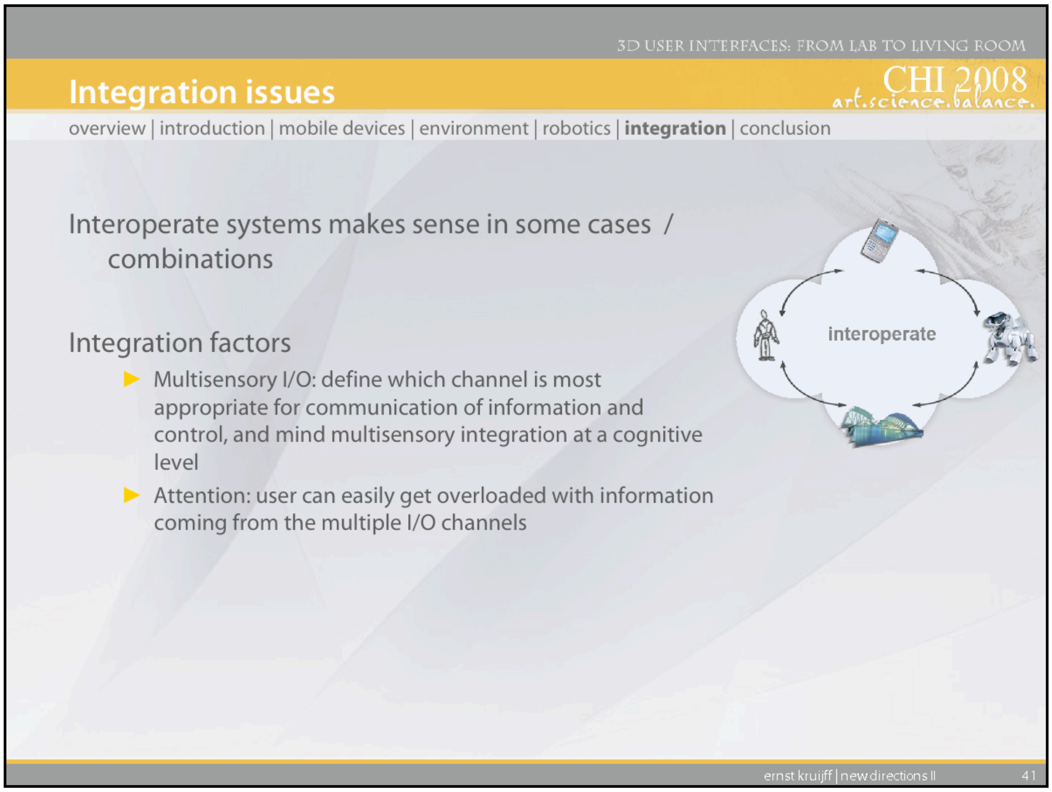
Recommended reading

Articles

- Fong et al. A survey of socially interactive robots. 2003
- Browse International Conference on Human-Robot Interaction (HRI) publications

Courses

- Scholtz, Yanco, Drury. Introduction to Human-Robot Interaction. CHI 2004 / IUI 2006.
- Kruijff, GJ. How to make talking robots? IEEE RO-MAN'07



In multiple cases, the previously mentioned fields can be combined. For example, mobile interfaces already exist for controlling robots. However, there are no „golden guidelines“ for integration, since the combinations of the different techniques may differ widely.

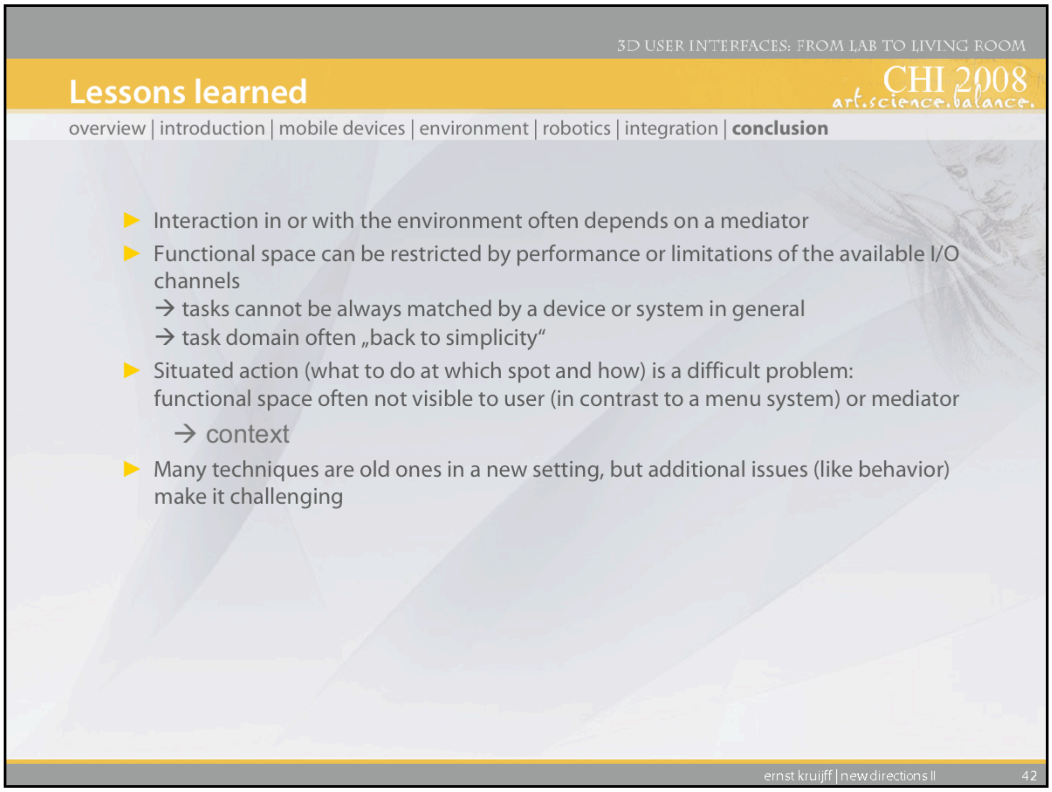
Nonetheless, some high-level integration tips can be given. First, take care of multisensory I/O and define which channels can be best used to convey specific information. The different sensory channels all have their specific strenghts and weaknesses: a wrong allocation of information can easily lead to cognitive overload, also due to possible cross-modal integration effects.

Related to cognitive overload: try to deal with attention and attentional resources. A user can easily get overloaded by information interpretation and/or the planning and performance of actions when multiple activities take place at the same time.

An excellent source that helps in many cases is Salvendy‘s handbook on human factors.

Reference

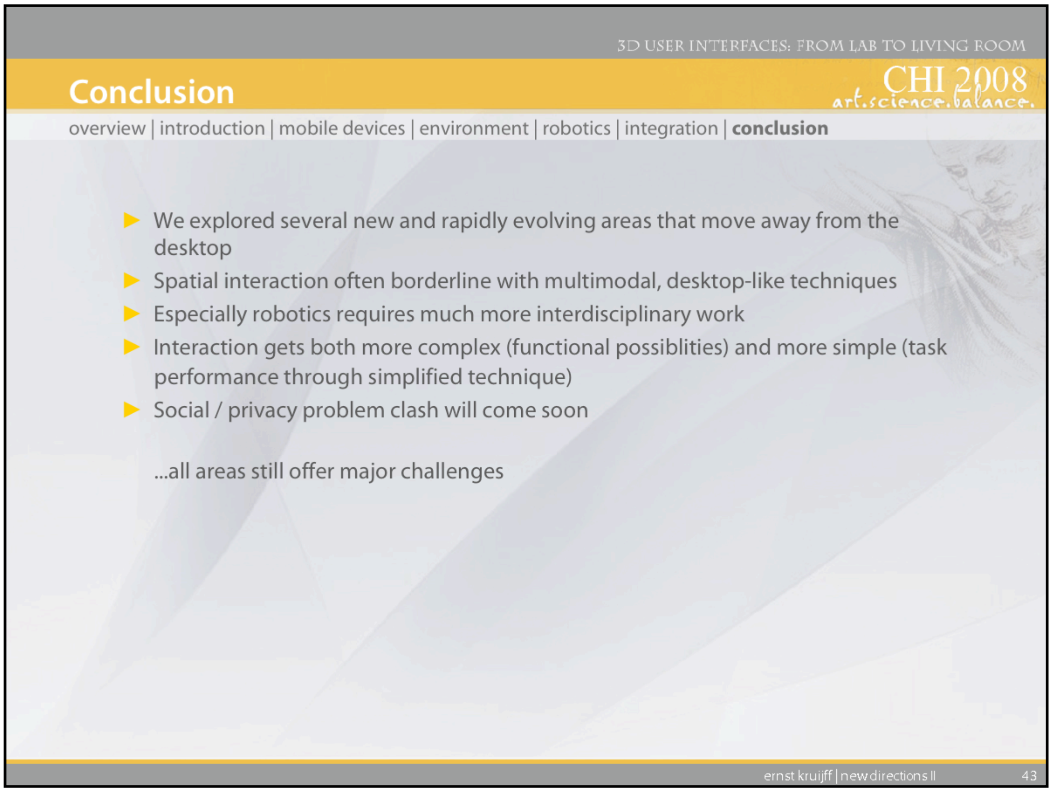
Salvendy (ed). Handbook of Human Factors and Ergonomics. John Wiley & Sons, 1997.



Reflecting the three different topics handled in this talk, what are the „lessons learned"?

First, it is often required to make use of a mediator to interact with an environment. Environmental displays are not widely available and will probably not be for a long time. As such, mobile devices or robots form an interesting alternative. Hereby, the functional space is limited most of the times, restricted by performance limitations of the I/O channels.

To deal with context / situation in so called „situated actions" stays a difficult problem. The functional space (possible tasks to be performed) is not always visible to either the user him/herself or the mediator. Finally, and not so surprising, most of the techniques presented in this talk are somehow direct deriviates of „traditional" techniques. Notwithstanding, innovation is high due to the additional issues that come into play, like behavior or emotion. There are still many open issues to deal with!



This brings us to the end of this talk. We have explored many fields that have evolved from or relate to „traditional“ 3D user interfaces. Nonetheless, as we could see in many examples, spatial interaction often was at the borderline with traditional desktop methods: there is still much potential for spatial interaction in all areas. Especially robotics requires considerable interdisciplinary work due to its high complexity and relation to many fields of research.

A strange phonemona one can also observe when looking at the tasks performed using spatial interaction: they get both more complex, functionally, and more simple, by ways of techniques applied. Finally, the social and privacy issues involved in many techniques presented will likely come to a clash soon. Currently, most of the associated issues are not solved.

Acknowledgements

- István Barakonyi
- Mark Billinghurst
- Geert-Jan Kruijff
- Erick Mendez
- Gerhard Schall
- Daniel Wagner


3D USER INTERFACES: FROM LAB TO LIVING ROOM
Coming up next...
CHI 2008
art.science.balance.
Video Games: 3DUIs for the Masses
Games and 3D interaction, consoles, arcades, pc games, …..
But first: Lunch break!
ernst kruijff | new directions II
45