Pattern Recognition as a Key Technology for the Next Generation of User Interfaces

Matthias Rauterberg and Patrick Steiger

Swiss Federal Institute of Technology (ETH)
Nelkenstrasse 11, CH-8092 ZUERICH, Switzerland
+41-1-632 70 82, rauterberg@ifap.bepr.ethz.ch

ABSTRACT

It is time to go beyond the established approaches in human-computer interaction. After a serious critic of command language, menu selection, and desktop interfaces we discuss the two known approaches to overcome the obstacles and limitations: [immersive] Virtual Reality (VR), and Augmented Reality (AR). Both design strategies are diametrically opposed: VR enriches the virtual world with real humans, while AR augments the real world with intelligent features. Only with the AR design strategy humans are able to behave as much as possible in a natural way: behavior of humans in the real world with other humans and/or real world objects. Our interest in human centred design let us follow this idea. Based on the fundamental constraints of natural way of interacting we derive a set of recommendations for the next generation of user interfaces: the *Natural User Interface* (NUI). The concept of NUI is discussed in form of a general framework and in form of several NUI-like applications. Finally we describe the interdisciplinary research topics that must be taken into consideration to come up one day--in the near future--with a well-designed NUI.

1. INTRODUCTION

The embodiment of computers in the work place has had a tremendous impact on the field of human-computer interaction. Mice and graphic displays are everywhere, the desktop workstations define the frontier between the computer world and the real world. We spend a lot of time and energy to transfer information between those two worlds. This could be reduced by better integrating the virtual world of the computer with the real world of the user. The most promising approach to this integration is *Augmented Reality* (AR) [44]. The expected success of this approach lies in its ability to build on fundamental human skills: namely, to interact with real world subjects and objects!

In this paper we present a new interaction technique, called Natural User Interface (NUI). Several prototypical applications already implement NUIs and demonstrate its technical feasibility.

2. WHY DO WE NEED A NEW INTERACTION TECHNIQUE?

Today several dialogue techniques are developed and in usage. The following dialogue techniques and dialogue objects can be distinguished with regard to traditional user interfaces: command language, function key, menu selection, icon, and window [37]. These techniques can be summarised into three different *interaction styles:*

- *Command language:* This interaction style (including action codes and softkeys) is one of the oldest way of interacting with a computer.
  Pros: In the command mode the user has a maximum of direct access to all available functions and operations.
  Cons: The user has no permanent feedback of all actual available function points.

- *Menu selection:* This includes rigid menu structures, pop-up and pull-down menus, form fill-in, etc. This interaction style became technically possible only with those terminals which, essentially, can reproduce only the ASCII character set. With this type of interaction style function keys are often used in addition to manage the dialogue.
  Pros: All available functions are represented by visible interaction points.
  Cons: Finding a function point in deeper menu hierarchies is cumbersome.

- *Direct manipulation*: This type of interaction style did not spread until bit mapped graphic displays came on the market; the development of this interaction style was based on the desktop metaphor which assumes that by depicting the work environment (i.e. of the desk: files, waste-paper basket, etc.) as realistically as possible on the I/O-interface, it would be particularly easy for the user to adjust to the virtual world of electronic objects.
  Pros: All functions are continuously represented by visible interaction points (e.g. mouse sensitive areas, etc.). The activation of intended functions can be achieved by directly pointing to their visible representations.
  Cons: Direct manipulation interfaces have difficulty handling variables, or distinguishing the depiction of an individual element from a representation of a set or class of elements.

In all these traditional interaction styles the user can not mix real world objects with virtual objects in the *same* interface space. They also do not take into considerations the enormous potential of human hands to interact with real and virtual world objects. This aspect was one of the basic ideas to develop data gloves and data suits for interactions in a virtual reality system (VR). The other basic idea, to realise a VR system, was the 3D output capabilities in the usage of head mounted displays. However, in VR systems several serious problems are inherently present; they are:
- The lack of tactile and touch information and consequently the mismatch with the proprioceptive feed

back. Special techniques are proposed to overcome this problem [10].

- The lack of information for depth perception, since visual displays generate 2D data. Many informational concepts are generating possibilities to reconstruct 3D pictures from these 2D data [32].

- There is always a time delay in the user-computer control loop, which may yield severe problems with reference to the perceptual stability of the vestibular apparatus in the ear [11].

- The strong influence of continuous communication--based on a shared social space--on social interaction is of paramount importance. Not only the shared sound space, but also the shared social nearness--in the real world--influences the human to human interaction [31]

The general advantage and disadvantage of immersive VR are the necessity to put the user into a complete modelled virtual world. This concept of immersing the user in the computer's world ignores the on-going process of interacting with the real world. In the same interface space the mixing of real and virtual objects is not possible. But, humans are--most of their time--part of a real world and interact with real objects and other real humans.

3. BEHAVIOR IN THE REAL WORLD

Each interaction with real world objects is constrained by the laws of physics (e.g., matter, energy, mechanics, heat, light, electricity, sound, etc.). More or less in a similar way, each interaction with real world humans is based on social and cultural norms. Social communication is multifunctional: One communication act (e.g., a message) encloses simultaneously different perspectives (e.g., referential, emotive, conative, relational, metalinguistically; see [17]). First, we describe necessary conditions for human to human communication. Second, we discuss aspects of human interactions with real world objects.

Human-human communication: One fundamental distinction was made by Watzlawick et al. [43]: Each human-human communication act has a *content* and a *relationship* part! Most of all existing technical information systems are dominated by the content part. The relationship part can be interpreted as a *metacontent level* to accompany or to control the content level of human-human communication (e.g., speech act).

Cassell and McNeill [5] [6] discovered that in face-to-face communication (e.g., storytelling), the status of content level is often indexed nonverbally, and is observed most clearly in the concurrent spontaneous gestures of the speaker. In many ways gestures add another dimension to the speech act--certain aspects of events may only be conveyed in gesture and not in speech, or vice-versa, or different aspects may be conveyed in each medium, giving the recipient a more complete view of the speaker's conception of the event.

McNeill [28] identified four different types of nonfacial gesture associated with speech: (1.) *Iconics* depict by the form of the gesture some feature of the action or event being described; such as "he climbed up the pipe"

accompanied by the hand rising upwards to show the path. (2.) *Metaphoric gestures*, which are also representational, but where the concept being depicted has no physical form. An example is "the meeting went on and on" accompanied by a hand indicating rolling motion. (3.) *Deictics*, which spatialize, or locate in the physical space in front of the narrator, aspects of the story being narrated; such as "Joe looked at Pattie, and Pattie looked back" accompanied by a hand pointing first to the left and then to the right. (4.) *Beat gestures*, small baton-like movements that do not change in form with the content of the accompanying speech. They serve a pragmatic function, occurring with comments on one's own linguistic contribution, speech repairs and reported speech. An example is "she talked first, I mean second" accompanied by a hand flicking down and then up.

These four gesture types are found in every culture, but universal is the types, not the shapes. Metaphorics are especially diverse, probably because metaphors themselves vary so widely from one language or culture to another. When speakers of English tell a story, for instance, they often begin with a "conduit" gesture--a handing over of information in the form of a package. Chinese speakers are more likely to spread out their hands, palms down, as if laying out a landscape. Overall, this quartet of nonfacial gestures plays specific roles with respect to the content level of the discourse they accompany. Iconic gestures occur at the content or narrative level, metaphoric and deictic gestures at the content and metacontent levels, and beat gestures only at the metacontent level. Most nonverbal gestures (including facial expressions) are the essence of the nonverbal communication part.

The content level of each verbal communication (spoken or written) transports facts, statements, messages and arguments (the *rational* part). The metacontent level refers to the semantic of gestures, body language, facial expressions and feeling aspects (the *emotional* part). The emotional aspects are often neglected and the rational aspects are overestimated. But, both levels are 'two sides of the same coin'! In human to human communication the best arguments can not be convincing, if the emotional level is lacking. This was one reason why internet users created smileys to improve their text based chatting sessions [36]. Smileys are a form of emotional icons, the so-called *emoticons* [34]. The empirical results of the study in [34] indicate that subjects using emoticons are more satisfied with the communication than those subjects not using emoticons. The outcomes of another empirical study showed the positive influence of emoticons on the quality of task results [33].

Additionally to the importance on nonfacial gestures we have to discuss the influence of facial expressions on human to human communication. The first concern of a review article of Fehr and Exlin [12] is with a conceptual orientation and it takes the view that humans' interaction is based "on an underlying shared code, the signs, symbols, or expressions of which have similar referents or meaning for code users" ([12], p. 228). The authors are aware of the problems attending the view of nonverbal behavior as embodying a set, or sets, of shared codes. They propose the position that the mean

ing of a nonverbal behavior derives, at least in part, from its use in one or more social contexts, its relation to those contexts and cultures of which they are part.

Fehr and Exlin [12] discuss visual interaction, particularly in terms of research concerned with the activity of the eyes, the orienting movements of the head and body, and their implications for the perceptions and behaviors of the interactants. The authors review studies that have to do with the occurrences of mutual gaze behavior and the functions it serves. Developmental differences in the use of, and response to, mutual gaze have been reported, as well as cross-cultural differences (e.g., gender differences appear very early in life, and visual attention of children appears to reflect the social organisation of which they are a part). The authors differentiate between (1) visual behavior in ongoing interaction, (2) gaze behavior to enduring characteristics of individuals, and (3) the relations of transient psychological states and cultural differences to gaze behavior.

The most important results are: (1) visual attention plays an important and, at times, critical communicative role in interpersonal encounters; (2) the information communicated frequently can be 'understood' only in terms of the context within which it occurs and the expectations of the recipient. That it is so constrained, however, does not make it less potent in its influence upon the course of an interaction and its participants' interpretations of each other's contributions. Direct and indirect gazes help to structure and regulate an interaction both to the extent that they operate to organise the alternation of the speakers and to the degree to which they offer and/or permit approaches and withdrawals on the parts of the interactants. Visual attention serves not only to impart information, it serves also as an information-gathering strategy and, as such, appears to be particularly crucial during infancy. Gaze behavior also appears to reflect, somewhat differently for males and females, the affective tone of an interaction (the *emotional* part). Finally, such behavior also appears to communicate certain personality characteristics of the participants in an interaction.

Human-object interaction: *Task related activities* can be discriminated into the following types: everyday skills of prehensile behavior [26], and other motor movements. The major classes of motor movements are: "(1) *Discrete movements* involve a single reaching movement to a stationary target, such as reaching for a control or pointing to a word on a computer screen; discrete movements can be made with or without visual control. (2) *Repetitive movements* involve a repetition of a single movement to a stationary target or targets. Examples include hammering a nail or tapping a cursor on a computer keyboard. (3) *Sequential movements* involve discrete movements to a number of stationary targets regularly or irregularly spaced. Examples include typewriting or reaching for parts in various stock bins. (4) *Continuous movements* involve movements that require muscular control adjustments of some degree during the movement, as in operating the steering wheel of a car or guiding a piece of wood through a band saw. (5) *Static positioning* consists of maintaining a specific position of a body member for a period of time. Strictly speaking, this is not a movement, but rather the absence of movement. Examples include

holding a part in one hand while soldering, or holding a needle to thread it" ([35], p. 277).

In the context of this paper we are primarily interested in purposeful motor activities, those that are executed by a person to achieve some goal (in delimitation from erroneous or exploratory behavior). Actions (e.g., based on motor movements) are functionally and not anatomically or mechanically specific. The catching of a ball could be carried out by either the left or the right hand, the starting position of the approach and the catching position of the ball might change from one reach to the next, and not two reaching trajectories will look exactly the same. These movements are still classified as the same action, though, because they share the same function.

Following the argumentation of Fitzmaurice, Ishii and Buxton [14] a graspable user interface has the following advantages:
"• It encourages two handed interactions;
• shifts to more specialised, context sensitive input devices;
• allows for more parallel input specification by the user, …
• leverages off of our well developed skills … for physical object manipulations;
• externalises traditionally internal computer representations;
• facilitates interactions by making interface elements more 'direct' and more 'manipulable' by using physical artifacts; …
• affords multi-person, collaborative use" ([14] p.443).

Summarising the above discussed results of real world behavior we can conclude the following design recommendations: (1) Each human to human communication consists of a *rational part* and an *emotional part*; both aspects must be integrated. (2) The distinction between content and metacontent level of communication leads to the separation of *working area* and *communication area*. (3) To empower the human to computer interaction the user must be able to behave in a *natural way*: bring into action all of his or her body parts (e.g., hands, arms, face, head, voice, etc.). To interpret all of these expressions we need very powerful and intelligent pattern recognition techniques.

4. A FRAMEWORK FOR NATURAL USER INTERFACES (NUI)

Augmented Reality (AR) recognises that people are used to the real world and that the real world cannot be reproduced completely and accurately enough on a computer. AR builds on the real world by augmenting it with computational capabilities. AR is the general design strategy behind "Natural User Interfaces" (NUI).

A system with a NUI supports the mix of real and virtual objects. As input it recognises (visually, acoustically or with other sensors) and understands physical objects and humans acting in a natural way (e.g., speech input, hand writing, etc.). Its output is based on pattern projection such as video projection, holography, speech synthesis or 3D audio patterns. A necessary condition in our definition of a NUI is that it allows inter-referential I/O [13], i.e. that the same mo

dality is used for input and output. For example, a projected item can be referred directly by the user for his or her nonverbal input behavior. Figure 1 provides an overview how a system with a NUI could look like.
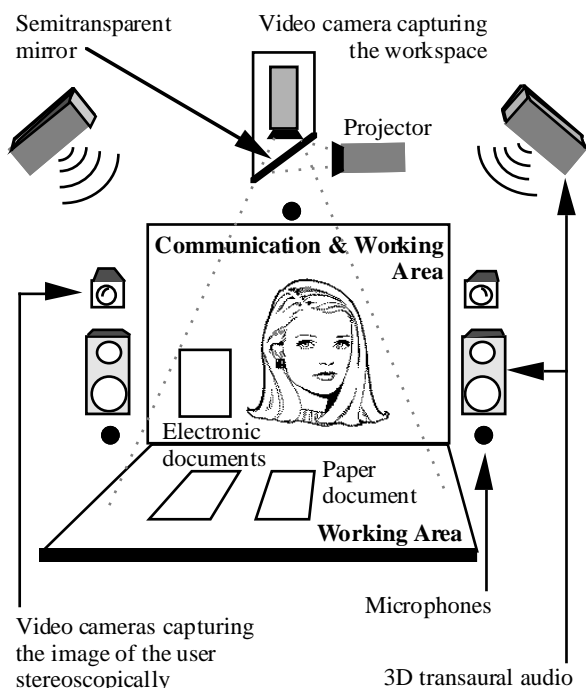

Figure 1: Components of a Natural User Interface

The spatial position of the user is monitored by two cameras. This also creates a stereoscopic picture for potential video conference partners. Speech and sound are recorded by several microphones, again allowing the system to maintain it's internal 3D model of the user. A third close-up camera on the top, records permanently the content state of the users work taking place on the horizontal working area. There, virtual and physical objects are fully integrated.

This set-up of several parallel input channels allows to show multiple views to remote communication partners, such as a (3D) face view [42] and a view of shared work objects [49]. Multimedia output is provided through the vertical communication area display, the projection device from the top down to the working area and through the four loudspeakers, producing a spatial impression on the user. Free space in the communication area can be used for (content) work, too (see Figure 1). Of course, traditional input and output devices still can be used in addition. As required by Tognazzini [39], NUIs are multimodal and therefore allow users to choose for every action the appropriate and individually preferred interaction style.

Since human beings manipulate objects in the physical world most often and most naturally with hands, there is a desire to apply these skills to human-computer interaction. In fact, NUIs allow the user to interact with real and virtual objects on the working area in a--literally--direct manipulative way! The working area is primarily horizontal, so that user can put real objects on the surface. Users get the feedback of the state of manipulated objects exactly at the same place where they

manipulate these objects: perception space and action space coincide!

5. 'NUI'-LIKE APPLICATIONS

There exist already several applications realising different NUI-like systems. These prototypes are:

• *DigitalDesk* [30]: Wellner developed the first Digital-Desk [45] [47]. It is based around an ordinary desk and can be used as such, but has extra capabilities. A video camera is mounted above the desk, pointing down at the work surface. The camera's output is fed through a pattern recognition system that can detect where the user is pointing and it can read documents that are placed on the desk. A computer-driven projector as an output device is also mounted above the desk, allowing the system to project electronic objects onto the work surface and onto real paper documents (something that can not be done with flat display panels or rear-projection).

• *Mosaic* [24]: This application is based on the Digital-Desk. Mosaic addresses the problem of extending paper along the temporal dimensions. Mosaic provides an interface that combines the benefits of paper storyboards with computer-controlled video.

• *Ariel* [25]: In a European ESPRIT project, the team of Mackay explored new ways of supporting engineers with a NUI-based system. They analysed the task of engineers constructing the world's longest suspension bridge in Denmark. In practice, engineers often work with physical paper copies of CAD-constructed engineering drawings. To update manually added annotations in the on-line version and to support communicating such information between engineers, physical drawings have been enriched with a digital media space. An engineer's personalised paper drawing becomes the interface to the computer system.

• *PlayingDesk* [33]: Rauterberg et al. realised a game to be played on a projected virtual playing field with *real* chips. First, the player puts a real black chip on a field. The system recognises the position, calculates the position of its next move and marks that field with a virtually projected red chip. In a field trial (during a 5-day fair) we had installed 4 stations, each offering the same game with a different user interface style (command interface, mouse-based GUI, GUI with touchscreen and the PlayingDesk/NUI). More than 8000 users played a game against one of our computers. We found a significant higher chance for the users to win if they played with the well-known game board (NUI, PlayingDesk) while the chance to win was smaller if they played at one of the other three stations with conventional user interfaces [33].

• *DoubleDigitalDesk* [46]: The DoubleDigitalDesk, an extended version of the DigitalDesk, makes it possible to share real paper documents. It allows both users in two separate locations to "share" their physical desks, for seeing, editing and writing on each other's paper documents.

• *InteractiveDesk* [2]: A computer image is displayed on the desk with an additional pen-input facility and an ordinary upright display with a keyboard, thus inte

grating conventional office systems and an augmented reality system. The prototype assists users in switching input method (based on location of the keyboard) and retrieving files using real objects.

• *BrightBoard* [38]: The BrightBoard concentrates on using video input to control a computer system. A camera records commands on a whiteboard, flip-chart or even a sheet of paper. Feedback is provided acoustically by sound or by a speech synthesiser. Though the BrightBoard is implemented based on the ideas of augmented reality, it is *not* a NUI in the definition given at the beginning of the previous section, since it does not support inter-referential I/O. The output is acoustical and the input is visual. The user has to learn a lot of commands (written on the board by hand) to control the system in the manner that is typical for a command user interface.

In all presented NUI-based applications the computer identifies the user's commands by video-based pattern recognition and executes the appropriate actions by projecting its output back into the same working area.

6. NUI-RELATED TECHNICAL RESEARCH AREAS

To realise a NUI, input and output have to be supported by pattern recognition techniques. The user interacts in his natural environment with real or virtual objects. The system captures the scene and processes the image to recognise (in the context) meaningful objects. This could be (1) information on a piece of paper or a whiteboard, it may be (2) a human-being acting on the scene, it could be (3) any flat or 3D object, or it can be (4) a digital object projected into the scene. Therefore, efficient and robust algorithms are required from the following research areas:

• 2D [3] and 3D [16] [21] [22] [41] object recognition,
• pattern recognition [7],
• scanning and optical character recognition [15],
• gesture recognition [18] [27] [40],
• voice [8] [50] and sound recognition [20],
• image understanding [9].

Is has to be determined which images are relevant and when the system has to take an action. E.g., if the user is moving his finger over a document to point on a particular word, and the system takes several images, how does it know on which image the user is arrived at his final position?

On the output side the following fields are involved:

• 2D projection,
• 3D projection [4] [19] [35],
• 3D audio rendering [48],
• speech [1] [8'] and sound [29] synthesis.

2D projection is a well-settled discipline, used in everyday situations, such as slide projection, TV projection, etc. 3D projection is still a research issue. Promising approaches are digital holography [23] and volumetric display systems [4]. The real-time integration of a 3D-object into a filmed (live) scene is described by Kansy et al. [19]. Requirements of such projections are flicker-free, high resolution images.

7. CONCLUSION

We have shown that Natural User Interfaces (NUI) have many advantages over traditional interaction styles and VR. Especially aspects of nonverbal communication between humans and capturing task-related activities (motor movements, voice and sound) are emphasised. Technology is already ripe enough that many research prototypes can show the feasibility of particular components of NUIs. Calling for multi-disciplinary research, we identified several domains we consider as key technologies to attack the remaining technical problems, which are mainly pattern recognition and projection. Providing people with support for their real-world tasks, NUIs open a wide new design space for the next generation of HCI technology.

8. REFERENCES

[1] Allen J, Hunnicutt S, Klatt D: From text to speech - the MITalk system. Cambridge Univ. Press, 1987.

[2] Arai T, Machii K, Kuzunuki S, Shojima H: InteractiveDESK: A Computer-augmented Desk Which Responds to Operations on Real Objects. In *Proc. of the CHI '95 Companion*, 1995, pp. 141-142.

[3] Bennamoun M, Boashash B, Koo J: Optimal Parameters for Edge Detection. In *Proc. of IEEE Intern. Conf. on Systems, Man and Cybernetics*, 1995, pp. 1482-1487.

[4] Blundell B, Schwarz A, Horrell D: Volumetric Three-dimensional Display Systems: Their Past, Present and Future. *Engineering Science and Education Journal*, October 1993, pp. 196-200.

[5] Cassell J, McNeill, D: Gesture and Ground. In *Proc. of the 16th AnnualMeeting of the Berkeley Linguistics Society, BLS* 16, 1990, pp. 57-68.

[6] Cassell J, McNeill D: Nonverbal Imagery and the Poetics of Prose. *Poetics Today* Vol. 12:3, 1991, pp. 375-404.

[7] Chandrasekaran V, Palaniswami M, Caelli T: Pattern Recognition by Topology Free Spatio-Temporal Feature Map. In *Proc. of IEEE Intern. Conf. on Systems, Man and Cybernetics*, 1995, pp. 1136-1141.

[8] Cohen P, Oviatt S: The role of voice input for human-machine communication. *Proc. Natl. Acad. Sci. USA,* 92(22), 1995, pp. 9921-9927

[8'] Cole R: Survey of the State of the Art in Human Language Technology. On URL: http: //www.cse .ogi.edu /CSLU /HLTsurvey /HLTsurvey.html

[9] Covell M: Autocorrespondence: Feature-based Match Estimation and Image Metamorphosis. In *Proc. of IEEE Intern. Conf. on Systems, Man and Cybernetics*, 1995, pp. 2736-2741.

[10] CyberTouch, Virtual Technologies Inc., 2175 Park Boulevard, Palo Alto, CA.

[11] DIVE Laboratories Inc: Health and HMDs. On URL: http://www.divelabs.com /deeper.html

[12] Fehr B, Exline R: Social visual interaction: a conceptual and literature review. In *Nonverbal behavior and commnication* (A. Siegman & S. Feldstein, eds.), Lawrence Erlbaum, 1987, pp. 225-326.

[13] Draper S: Display Managers as the Basis for User-Machine Communication. In Norman D, Draper S (eds.) *User Centered System Design*. Lawrence Erlbaum, 1986, pp. 339-352.

[14] Fitzmaurice G, Ishii H, Buxton W: Bricks: Laying the Foundations for Graspable User Interfaces. In *Proc. of the CHI ´95*, 1995, pp. 442-449.

[15] Govindaraju V, Srihari S: System for Reading Handwritten Documents. In *Proc. of IEEE Intern. Conf. on Systems, Man and Cybernetics*, 1995, pp. 2347-2352.

[16] Hara K, Zha H, Nagata T: Curve and Surface Reconstruction by Using Sequential Markov Random Fields (MRFs). In *Proc. of IEEE Intern. Conf. on Systems, Man and Cybernetics*, 1995, pp. 1097-1102.

[17] Jakobson R: Closing statement: linguistics and poetics. In *Style in language* (T. A. Sebeok, ed.) Cambridge, 1960, pp. 350-377.

[18] Kämmerer B, Maggioni C: GestureComputer - Research and Practice. In *Proc. of 4th Intern. Conf. INTERFACE to Real & Virtual Worlds*, 1995, pp.251-260.

[19] Kansy K, Berlage T, Schmitgen G, Wisskirchen P: Real-Time Integration of Synthetic Computer Graphics into Live Video Scenes. In *Proc. of 4th Intern. Conference INTERFACE to Real & Virtual Worlds*, 1995, pp. 93-101.

[20] Kramer G (ed.): Auditory Display: Sonification, Audification and Auditory Interfaces. *SFI Studies in the Sciences of Complexity, Proc. Vol. XVIII*, Addison-Wesley, 1994.

[21] Li Z, Chin H: Depth and Occlusion Recovery in Motion Stereo. In *Proc. of IEEE Intern. Conf. on Systems, Man and Cybernetics*, 1995, pp. 3890-3895.

[22] Liu R, Durdle N, Peterson A: Three Dimensional Reconstruction of Trunk Surface Using Structured Light. In *Proc. of IEEE Intern. Conf. on Systems, Man and Cybernetics*, 1995, pp. 1085-1090.

[23] Lucente M, Galyean T: Rendering Interactive Holographic Images. In *Proc. of SIGGRAPH'95*, 1995, pp. 387-394.

[24] Mackay W, Velay G, Carter K, Ma C, Pagani D: Augmented Reality - Adding Computational Dimensions to Paper. *Communications of the ACM*, 36(7), 1993, pp. 96-97.

[25] Mackay W, Pagani D, Faber L, Inwood B, Launiainen L, Brenta L, Pouzol V: Ariel: Augmenting Paper Engineering Drawings. In *Proc. of the CHI ´95 Companion*, 1995, pp. 421-422.

[26] MacKenzie C, Iberall T: The grasping hand. Elsevier, 1994.

[27] Masaaki F, Kenji M, Yasuhito S: A Glove-Free Interface by Hand Image Processing. In *Proc. of imagina ´93*, 1993, pp. 122-126.

[28] McNeill D: Hand and Mind: What gestures reveal about thought. University of Chicago Press, 1992.

[29] Munteanu E, Guggiana V, Darvishi A, Schauer H, Rauterberg M, Motavalli M: Physical modelling of environmental sounds. In: *Proc. of '2nd Intern. Conf. on Acoustic and Musical Research'–CIARM' 95* (F. Pedrielli, ed.) Ferrara, 1995, pp. 107-112.

[30] Newman W, Wellner P: A Desk Supporting Computer-base Interaction with Paper Documents. In *Proc. of the CHI ´92*, 1992, pp. 587-592.

[31] Rauterberg M, Dätwyler M, Sperisen M: From Competition to Collaboration through a Shared Social Space. In: *Proc. of East–West Intern. Conf. on Human–Computer Interaction (EWHCI '95)*, 1995, pp. 94-101.

[32] Rauterberg M, Szabo K: A Design Concept for N-dimensional User Interfaces. In *Proc. of 4th Intern. Conf. INTERFACE to Real & Virtual Worlds*, 1995, pp. 467-477.

[33] Rauterberg M, Stebler R, Mauch T: "Augmented Reality" in Contest with a Command, Desktop and Touch Screen Interface. In *Proc. of 5th Intern. Conf. INTERFACE to Real & Virtual Worlds*, 1996, pp. 95-103.

[34] Rivera K, Cooke N, Bauhs J: The effects of emotional icons on remote communication. In *Proc. of the CHI ´96 Companion*, 1996, pp. 99-100.

[35] Sanders M, McCormick E: Human Factors in Engineering and Design. McGraw Hill, 1993.

[36] Sanderson D: Smileys. O'Reilly & Assoc., 1993.

[37] Shneiderman B: Designing the User Interface. Addison-Wesley, Reading MA, 1987.

[38] Stafford-Fraser Q, Robinson P: BrightBoard: A Video-Augmented Environment. In *Proc. of the CHI ´96*, 1996, pp. 134-141.

[39] Tognazzini B: *Tog on Software Design*. Addison-Wesley, Reading MA, 1996.

[40] Voyles R, Khosla P: Multi-Agent Gesture Interpretation for Robotic Cable Harnessing. In *Proc. of IEEE Intern. Conf. on Systems, Man and Cybernetics*, 1995, pp. 1113-1118.

[41] Wang Z, Ohnishi N: Deformable Template Based Stereo. In *Proc. of IEEE Intern. Conf. on Systems, Man and Cybernetics*, 1995, pp. 3884-3889.

[42] Watts L, Monk A: Remote assistance: a view of the work and a view of the face?. In *Proc. of the CHI ´96 Companion*, 1996, pp. 101-102.

[43] Watzlawik P, Beavin J, Jackson D: Pragmatics of Human Communication. A Study of Interactional Patterns, Pathologies, and Paradoxes. Norton & Company, New York, 1967.

[44] Wellner P, Mackay W, Gold R: Computer-Augmented Environments: Back to the Real World. *Communications of the ACM*, 36(7), 1993, pp. 24-26.

[45] Wellner P: Interacting with Paper on the Digital Desk. *Communications of the ACM*, 36(7), 1993, pp. 87-96.

[46] Wellner P, Freeman S: The DoubleDigitalDesk: Shared Editing of Paper Documents. *Technical Report EPC-93-108, Xerox Research Centre Cambridge Laboratory*, Cambridge UK, 1993.

[47] Wellner P: Interacting with Paper on the Digital-Desk. *Technical Report 330 (PhD Dissertation)*, University of Cambridge Computer Laboratory, Cambridge UK, March 1994.

[48] Wenzel E: Localization in Virtual Acoustic Displays. *Presence* 1(1), 1992, pp. 80-107.

[49] Whittaker S: Rethinking video as a technology for interpersonal communications: theory and design implications. *Intern. Journal of Human-Computer Studies*, 42, 1995, pp. 501-529.

[50] Zue V et al: Spoken Language Systems for Human/Machine Interfaces. In *Intelligent Text and Information Handling RIAO91*. (A. Lichnerowicz, ed.) Elsevier, 1991, pp. 936-955.

1996 IEEE International Conference on Systems, Man and Cybernetics

Information Intelligence and Systems

Beijing, China
October 14-17, 1996

Volume 4 of 4
96CH35929   **Sponsored by Tsinghua University**