Fjeld M., Bichsel M., Rauterberg M. (1998) BUILD-IT: An intuitive design tool based on direct object manipulation. In: Wachsmuth I., Fröhlich M. (eds) Gesture and Sign Language in Human-Computer Interaction - GW 1997. (Lecture Notes in Computer Science, vol 1371, pp. 297-308) Springer, Berlin, Heidelberg. https://doi.org/10.1007/BFb0053008

BUILD-IT: An Intuitive Design Tool Based on Direct Object Manipulation

Morten Fjeld[1], Martin Bichsel[2] & Matthias Rauterberg[1]

[1]Institute for Hygiene and Applied Physiology (IHA)
fjeld@iha.bepr.ethz.ch, rauterberg@iha.bepr.ethz.ch
[2]Institute of Construction and Design Methods (IKB)
mbichsel@ikb.mavt.ethz.ch
Swiss Federal Institute of Technology (ETH)
Clausiusstrasse 25, CH-8092 Zurich, SWITZERLAND
www.iha.bepr.ethz.ch/pages/leute/fjeld/fjeld.htm

Natural interaction, in the context of this paper, means human action in a world of tangible objects and live subjects. We introduce the concept of action regulation and relate it to observable human behaviour. A tool bringing together motor and cognitive action is a promising way to assure complete task regulation. Aiming for such tools, we propose a set of guidelines for the next generation of user interfaces, the *Natural User Interface* (NUI). We present a NUI instantiation called BUILD-IT, featuring video-mediated interaction in a task specific context. This multi-brick interaction tool renders virtual objects tangible and allows multiple user simultaneous interaction in one common space. A few user experiences are briefly described.

Keywords: Augmented Reality, natural interaction, Natural User Interface, graspable objects, computer mediated design

1 Virtual and Augmented Reality

The introduction of computers in the work place has had a tremendous impact on task solving methods. Mouse based and graphical displays are everywhere; desktop workstations define the frontier between digital (computer) and analogue (*real*) worlds. Time and energy is spent transferring information between such worlds. This effort could be reduced by better integration of the *virtual* computer world with the *real* user world and vice versa.

Several dialogue techniques were developed in the past, which are now in use. The following dialogue techniques and objects can be distinguished: command language, function key, menu selection, iconic and window [20]. These five essential terms can be cast into three different interaction styles: command language, menu selection and direct manipulation. They all have in common that the user cannot combine *real* world and *virtual* objects within the *same* interface space. Nor do they adequately incorporate the human hands enormous potential for interaction with real and virtual objects. These drawbacks gave reason to develop data gloves and data suits. Users equipped with such artefacts can more easily interact in an immersive, Virtual Reality (VR) system. Another reason to realise VR systems was the emergence of head mounted displays with 3D output capabilities. However, VR systems are still subject to serious, inherent limitations such as:

- Lack of tactile and touch information, giving a mismatch with the proprioceptive feedback. Special techniques are proposed to overcome this problem [6].
- Delay in the user-computer control loop, often yielding severe problems with reference to the perceptual stability of the ear vestibular apparatus.
- Interference between electronic communication and social interaction. A shared sound space and a shared social world can bring remedy and stimulate humans to mutual interaction [15].

The advantage, but at the same time disadvantage, of immersive VR is the necessity to put users into a fully modelled, intangible world. Users are most often part of a world where they interact with material objects and live humans. This on-going, *real* world interaction is ignored by modelled worlds, since mixing of tangible and virtual objects is not yet possible.

To overcome these drawbacks of immersive VR, the concept of *Augmented Reality* (AR) [23] was introduced. This approach is promising because it incorporates fundamental human skills: interaction with real world subjects and objects. Hence, the AR design strategy enables humans to behave in a nearly natural way. We call this way of behaviour *natural (inter-) action.*

In the following section, we elaborate on a concept for natural interaction. Based on this concept we derive design guidelines for the next generation of user interfaces, *Natural User Interfaces* (NUIs). Based on these guidelines and the existing AR approach, section 3 suggests a general NUI framework. Section 4 describes a task specific instantiation of the NUI framework, called BUILD-IT. This tool features video-based interaction, supporting construction and plant design. Section 5 offers a reference to some user experiences.

2 Natural Interaction

2.1 Observable action

For some years, interaction with machines has become part of everyday life. According to Winograd and Flores [25] "reflection and abstraction are important, but not the basis for everyday actions". We are interested in reflection, dialogue and communication when they are cause and/or effect of physical behaviour. Therefore, the focus of this article will be activity that is based on observable task solving behaviour. Further support is offered by Campbell [5], who found that the more a learning process includes task related actions, the greater the retention of this learning, and, the greater the transfer to new situations. According to Campbell's experiments [5], decision must be tied to action in order to be efficient, at least in the early stages of learning. After this clarification, we go ahead to see what observable behaviour may look like.

Describing body action in general was undertaken by Sanders and McCormick [19] with their classes of motor behaviour: *discrete*, *repetitive*, *sequential* and *continuous movement* as well as *static positioning*. Focusing on hand movements, interesting work was offered by Mackenzie and Iberall [13], who outlined *prehension behaviour* as "the application of functionally effective forces by the hand to an object for a task, given numerous constraints." Prehension actually has a double meaning: taking hold of with the *hand* and *mental* apprehension. We take this duality as a pretext to make *reflection* part of our discussion.

2.2 Motor and cognitive action

Aicher [3] shows that the relation between reflection and body is so close that cognitive processing is often rendered visible by the language of the hands. This would mean that mental activity has a strong relation with manual activity and vice versa. Aicher describes regulation of human activity as a cycle of action, comparison and correction, leading to new action. He claims that regulation of human activity is often reduced to inner, rational activity. So, the manual part of the cycle has lost its right importance. This lack of *doing* in human problem solving may be due to the sharp division between mental and manual work. Cognitive processing excludes observable action and vice versa. Hence, there is a need for a unifying concept within human computer interaction bringing together action and reflection.

2.3 Action regulation theory

Action regulation theory [10] is one well founded answer to this need. Based on task analysis, this tradition unifies action and reflection. Great importance is attached to the concept of the *complete task*. Each complete activity cycle starts with a goal setting part. The characteristics of a complete task are given by four distinctive steps (Fig. 1):

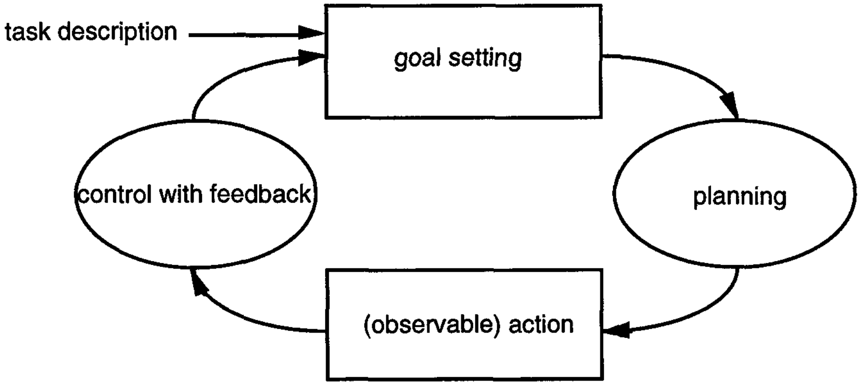

Fig. 1: A complete activity cycle in the context of Action Regulation Theory [10].

- *Independent setting of (sub-)goals, embedded in the superimposed task goal.*
- *Independent action preparation in the sense of taking on planning functions, and selection of the tools including the necessary actions for goal attainment.*
- *Physical (or event mental) performance functions with feedback on performance pertaining to possible corrections of actions.*
- *Control with feedback on results and the possibility of checking the results of one's own actions against the set (sub-)goals.*

Based on action regulation theory, we set out to enlarge the view of goal oriented, motor activity.

2.4 Epistemic and pragmatic action

Motor activity was classified by Kirsh and Maglio [12] as being either epistemic or pragmatic. Pragmatic actions have the primary function of bringing the user physically closer to a goal. In contrast, epistemic actions are chosen to unveil hidden information or to gain insight that otherwise would require much mental computation. Hence, physical action facilitates mental activity, making it faster and more reliable. Also, cognitive complexity may be reduced by epistemic actions. For instance, the memory involved in mental computation can be reduced or the number of steps required made fewer. As a consequence, the probability of errors due to mental slips is reduced.

The distinction of epistemic and pragmatic action was also mentioned by Gibson [9], suggesting that hand movements can be classified as *exploratory* and *performatory*. A similar distinction was made by Rotenstreich [18], who considered human activity to be either *play* or *labour*.

2.5 NUI design guidelines

At this point we are led to the question: can epistemic action be considered to be a goal-driven, task related activity? The answer to this question depends on the level of abstraction. For instance, epistemic action might be goal oriented when the aim is to learn and internalise a tool. Such a process is described by Kaptelinin [11], through the following steps:

- *The initial phase, when performance is the same with and without a tool because the tool is not mastered well enough to provide any benefits,*
- *the intermediate stage, when aided performance is superior to unaided performance, and*
- *a final stage, when performance is the same with and without the tool but now because the tool-mediated activity is internalised and the external tool (such as a checklist or a visualisation of complex data) is no longer needed.*

This three step process may be classed as epistemic action, but at the same time it constitutes task related, purposeful activity. So, epistemic action can certainly be considered to be task oriented. As a result, we find it worth-while to formulate the first point of our design guidelines: NUIs should allow for epistemic as well as pragmatic action.

Since exploring, feedback, comparison and correction all are ways to increase knowledge, we find it in focus to refer Buckminster Fuller's work on education. He showed how learning is intimately related to making mistakes. His idea is that "the more mistakes the students discover, the higher their grade" [4]. His proposal was actually to appreciate regulation of human activity. Students with a good sense for goal setting, planning, action and feedback get higher grades. We use this idea as a second point for our guidelines: Users should be allowed to behave in an exploratory way and to make mistakes. Such behaviour should only imply low risk and shall give constructive feedback.

To reach complete task solving in computer mediated activity we believe that users must behave in a *natural* way. They must be able to employ everyday motor faculties by bringing into action all of their body parts like hands, arms, face and voice. That requirement constitutes the third point of our design guidelines. Consequently, NUIs must be able to interpret a range of human expressions. So, they will need powerful and intelligent pattern recognition methods [16].

Summing up this section, we get the following set of guidelines for NUI design:

- Allow epistemic as well as pragmatic action.
- Assure that mistakes only imply low risk so that exploratory behaviour is being stimulated.
- Allow users to employ everyday gestures and motor patterns using all of their body parts like hands, arms, face and voice.

3 A general framework

In this section we make use of the elaborated guidelines for NUI design by extending the concept of Augmented Reality (AR). AR recognises that people are used to the *real* world, which cannot be authentically reproduced by a computer. AR is based on the *real objects*, augmented by computer characteristics. We take it as the general design strategy behind NUIs [17].

NUIs support the fusion of real and virtual objects. Being multi-modal interfaces, they understand visual, acoustic and other human ways of expression. They recognise physical objects and human actions like speech and hand writing in a natural way. Their output is based on techniques like video, holography, synthesised speech and spatial sound. NUIs necessarily imply inter-referential input/output [7], meaning that the same modality is used for input and output. Hence, a projected item can be referred to directly by users as part of their non-verbal expressions. Fig. 2 shows the architecture of one possible NUI instantiation.

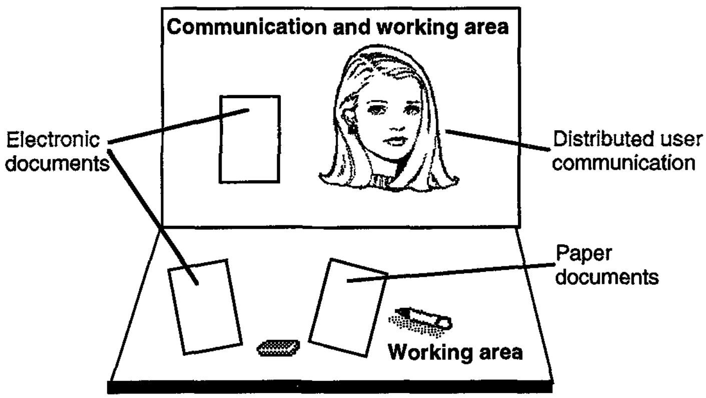

Fig. 2: Architecture of a Natural User Interface.

The spatial position of the user is monitored by one or more cameras. This could also create a stereoscopic picture for potential video conference partners. Speech and sound is recorded by several microphones, enabling the system to maintain an internal, spatial user model. From above, a close-up camera permanently records the state of the user activity taking place in the horizontal working area. In this very area, virtual and physical objects are fully integrated.

The use of several parallel input channels makes it possible to communicate multiple views to remote partners, such as 3D face images [22] and shared work objects [24]. Multimedia output as shown in Fig. 2 is provided by a) a vertical display for the communication and working area, b) a projection device illuminating the horizontal working area, and c) a multichannel audio system. Of course, traditional input/output devices can be added. As required by Tognazzini [21], NUIs are multimodal, so users are allowed to (re-)choose their personal and appropriate interaction style at any moment.

Since humans often and easily manipulate objects in the real world with their hands, they have a natural desire to employ this faculty when interacting with computers. NUIs allow users to interact with real and virtual object in a *literally* direct manipulative way. A planar working area allows them to place real objects onto a surface. There is a direct mapping of the real, user manipulated object onto its corresponding virtual object. We can actually say that perception and action space co-incide. This is a powerful design principle, empirically validated by Rauterberg [14].

4 A NUI prototype: BUILD-IT

In a first step, we have designed a weak NUI instantiation. *Weak* means that distributed communication remains to be implied. As task context, we chose that of planning activities for plant design. A prototype system, called BUILD-IT, was realised. This is an application that supports engineers in designing assembly lines and building plants.

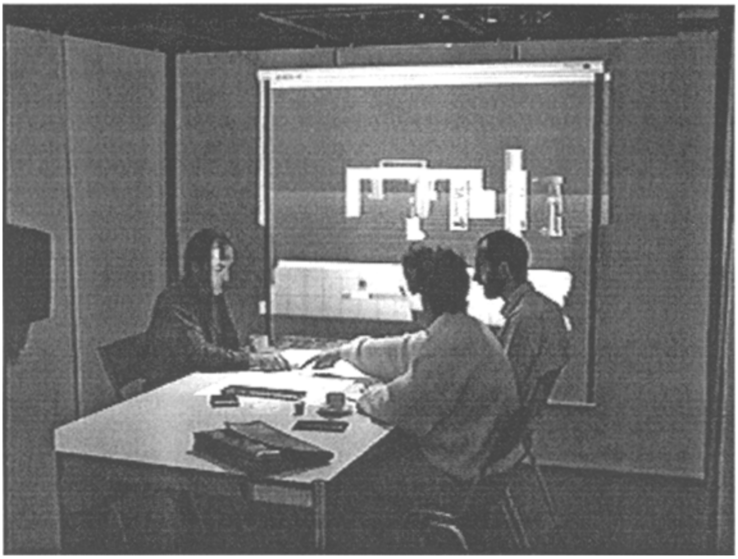

Fig. 3: The design room of BUILD-IT.

The design room of Fig. 3 enables users, grouped around a table, to interact in a space of virtual and real world objects. The vertical working area in the background gives a side view of the plant. In the horizontal working area there are several views where objects can be selected and manipulated.

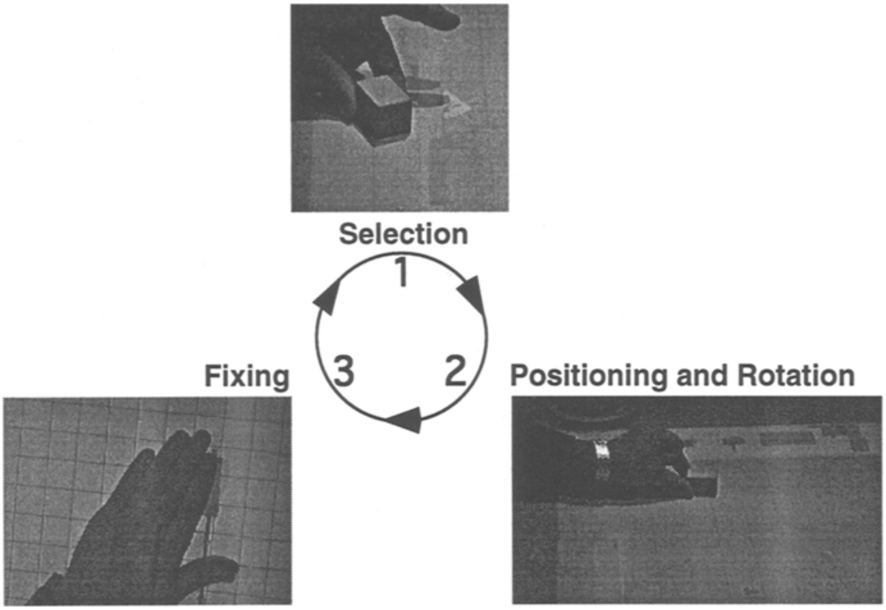

Fig. 4: The basic steps for user manipulations with the interaction handler.

The working principle of BUILD-IT is shown in Fig. 4. The user selects the object by putting the brick at the object positions. The object can be positioned, rotated and fixed by simple brick manipulation. Using a material brick, everyday motor patterns like grasping, moving, rotating and covering are activated. Throughout these steps, there is a strong connection between cognitive processing and observable behaviour. The system dynamically supports the user needs for goal setting, planning, action and control. Hence, complete regulation of the working cycle is assured. The cost of making a mistake is low, since all vital operations are reversible. So, epistemic and pragmatic action are equally supported. To allow two handed operation, the system supports multi-brick interaction. A second effect of multi-brick interaction, is that several users can take part in a simultaneous design process. Altogether, the set of guidelines we formulated in section 2 have been met.

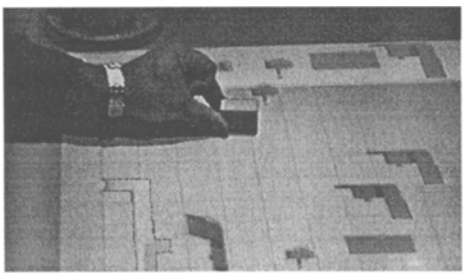

Fig. 5: The object menu (white), above view (grey) and interaction handler (brick).

The application is designed to support providers of assembly lines and plants in the early design processes. Graphical display is based on the class library MET++ [1]. The system can read and render arbitrary virtual 3D objects as seen in Fig. 5. These objects are sent from a Computer Aided Design (CAD) system to BUILD-IT using Virtual Reality Modelling Language (VRML).

Geometry is not the only aspect of product data. There is a growing need to interact in other dimensions, such as cost, configurations and variants. Therefore, the system has been engineered to send and receive numerous forms of metadata.

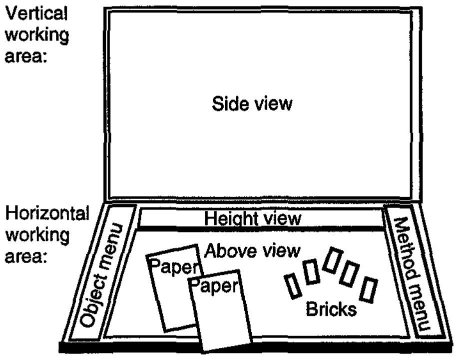

Fig. 6: The two working areas and their views.

BUILD-IT currently features the following user (inter-) actions (Fig. 4-6):

- Selection of a virtual object (e.g. a specific machine) in a virtual machine store by placing the interaction handler onto the projected image of the machine in the object menu.
- Positioning of a machine in the virtual plant by moving the interaction handler to the preferred position in the above view of the plant layout.
- Rotation of a machine by coupling machine and brick orientation.
- Fixing the machine by manually covering the surface of the interaction handler and then removing it.
- Re-selection of a machine by placing the interaction handler onto the specific machine in the above view.
- Removing the machine by moving it back into the object menu (the virtual machine store).
- Modification of object size and height by operators in the method menu applied on objects in the above view.
- Direct modification of object altitude in the height view.
- Automatic docking of two or more objects along predefined contact lines within the above view.
- Scrolling of above view, height view and menus.
- Modification of the perspective in the height and side views by cameras manipulation in the above view. Numerous cameras, each representing a distinct

perspective, can exist at a time. The last one selected determines the current perspective.

- Saving of the working area contents by a method menu icon.
- Printing of the views, also offered by a method menu icon.
- Multi-brick and multi-person interaction. All the previous (inter-) actions can be simultaneously executed by any of the bricks at the table.
- Simulation mode, supported by a simulation software [2], shows real-time manufacturing. Steel sheets can be followed as they pass through different processes, like laser welding, chemical baths and drilling.

5 User experiences

As system designers, we carried out various talks around the table (Fig. 3). Therefore, we consider ourselves to be the first users of BUILD-IT. This novel swapping of roles between *designer* and *user*, proved to be a stimulating one. Introducing and working out aspects like real and virtual objects, various view layouts as well as animation, was all mediated by those talks around the table. Some of these talks would not have been conceivable if we had worked in a conventional, screen-based setting.

The BUILD-IT system was also tried out with managers and engineers from companies producing assembly lines and plants. These tests showed that the system is intuitive and enjoyable to use as well as easy to learn. Most people were able to assemble virtual plants after only half a minute of introduction. Some typical user comments were: "The concept phase is especially important in plant design since the customer must be involved in a direct manner. Often, partners using different languages sit at the same table. This novel interaction technique will be a means for completing this phase efficiently.", "This is a general improvement of the interface to the customer, in the offering phase as well as during the project, especially in simultaneous engineering projects.", and "The use of this novel interaction technique will lead to simplification, acceleration and reduction of the iterative steps in the start-up and concept phase of a plant construction project".

6 Conclusion and future perspectives

One of the most interesting benefits of a NUI-based interface is the possibility to combine real and virtual objects in the same interaction space. Taking this advantage even further, we have implemented a framework for multiple interaction handlers, allowing simultaneous interaction of several users grouped at one single table. This possibility calls for detailed research into bimanual gestures and multi-person interaction. In particular, we want to look at the concept of time- and space-multiplexing, as described by Fitzmaurice [8].

Focusing at system performance within the task domain, we have observed how novice users compare with experts. We noticed that all customers whether CAD experts or not, could take part in discussions and management of complex 3D objects. Products and technical descriptions can easily be presented, and new requirements are realised and displayed within short time. The virtual camera allows a walk-through of the designed plant. Such inspection tours can give valuable input about complex systems.

In the near future, one could imagine a direct, NUI-based information flow between customers and large product databases. It is conceivable that planners wanting to change one detail of a machine, will have several configurations presented on their table. As soon as one has been selected, the exact configuration cost will be calculated and displayed.

7 References

[1] Ackermann P: Developing Object-Oriented Multimedia Software Based on the MET++ Application Framework. Heidelberg: dpunkt Verlag für digitale Technologie, 1996.

[2] AESOP GmbH: SIMPLE++ for modelling, simulation and visualisation, 1997.

[3] Aicher O: analog und digital, ernst & sohn, 1991, pp 19-21, 28.

[4] Buckminster Fuller R: On Education, The University of Massachusetts Press, Amherst, 1979, p. 46.

[5] Campbell J: Training design for performance improvement. In J. Campbell & R. Campbell (Eds.), Productivity in organisations (pp. 177-216). San Francisco: Jossey-Bass, 1988.

[6] CyberTouch, Virtual Technologies Inc., 2175 Park Boulevard, Palo Alto, CA 94306, http://www.virtex.com/

[7] 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.

[8] Fitzmaurice G W: Graspable User Interfaces, PhD at the University of Toronto, 1996. http://www.dgp.utoronto.ca/people/GeorgeFitzmaurice/thesis/Thesis.gf.html.

[9] Gibson, J. J. (1962). Observations on active touch. Psychological Review, 69, 477-491.

[10] Hacker W: Action regulation theory and occupational psychology. Review of German empirical research since 1987. The German Journal of Psychology 18(2) (1994) pp. 91-120.

[11] Kaptelinin V: Computer-Mediated Activity: Functional Organs in Social and Developmental Contexts. In Nardi B (ed.) Context and Consciousness, MIT Press 1996, p. 62.

[12] Kirsh, D. and Maglio, P. (1994). On Distinguishing Epistemic from Pragmatic Action, *Cognitive Science, 18, 513-549.*

[13] Mackenzie C, Iberall T: The grasping hand. Elsevier, 1994, p. 15.

[14] Rauterberg M., Über die Quantifizierung software-ergonomischer Richtlinien. PhD Thesis, University of Zurich, 1995.

[15] 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.

[16] Rauterberg M, Fjeld M, Krueger H, Bichsel M, Leonhard U, Meier M: BUILD-IT: A computer Vision-based Interaction Technique for a Planning Tool. In (Thimbleby H,

O'Connaill B, Thomas P eds.) People and Computers XII, Proc. of HCI'97, Springer, 1997, pp. 303 - 314.
[17] Rauterberg M, Steiger P: Pattern recognition as a key technology for the next generation of user interfaces. In Proc. of IEEE International Conference on Systems, Man and Cybernetics--SMC'96 (Vol. 4, IEEE Catalogue Number: 96CH35929, pp. 2805-2810). Piscataway: IEEE.
[18] Rotenstreich N: Reflection and Action. Martinus Nijhoff Publishers, 1985, p. 62.
[19] Sanders M, McCormick E: Human Factors in Engineering and Design. McGraw Hill, 1993, p. 277.
[20] Shneiderman B: Designing the User Interface. Addison-Wesley, Reading MA, 1987.
[21] Tognazzini B: Tog on Software Design. Addison-Wesley, Reading MA, 1996.
[22] 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.
[23] Wellner P, Mackay W, Gold R: Computer-Augmented Environments: Back to the Real World. Communications of the ACM, 36(7), 1993, pp. 24-26.
[24] 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.
[25] Winograd T, Flores C F: Understanding computers and cognition: a new foundation for design. Ablex publishing corporation, 1986, p. 97.