Continuous Natural User Interface: Reducing the Gap Between Real and Digital World

Nils Petersen *
DFKI Kaiserslautern

Didier Stricker †
DFKI Kaiserslautern

Abstract

Augmented reality (AR) presentation enables the creation of natural user interfaces that employ the whole user's environment as interaction device. Additionally, by using hand based 3D interaction with gestures that have a physical meaning like grabbing, dragging, and dropping this leads to a user experience that is intuitive, since close to the real world's behavior. We propose a novel approach to an AR-based natural user interface, that goes one step further by enabling the contents of the interface to switch domains from a virtual instance in AR to a physical instance in the real-world. All instances stay associated and changes made to the physical instance will be reflected on the virtual one. Because the behavior of our interface in AR is in key aspects consistent with the real-world, the gap between those domains is made less salient.

To demonstrate our concept, we have implemented an exemplary industrial use case. Our main contribution is the methodology for an intuitive interface we call continuous natural user interface (CNUI). Additionaly, we conducted a user study to investigate the acceptance of this kind of interface. Results indicate an ergonomic ease and after a training period also an increased performance when using our system.

Index Terms: H.5.2 [INFORMATION INTERFACES AND PRESENTATION]: User Interfaces—Input devices and strategies; H.5.1 [INFORMATION INTERFACES AND PRESENTATION]: Multimedia Information Systems—Artificial, augmented, and virtual realities; I.3.6 [COMPUTER GRAPHICS]: Methodology and Techniques—Interaction techniques

1 Introduction

The interest in novel user interfaces, that do not or not solely deal with mouse and keyboard, has increased recently. This can be backed by the emergence of and increased interest in multi touch tables [14], multi touch mobile devices [2], and arm-gesture controlled gaming consoles, etc., summarized under the term natural user interface (NUI).

The reason, why e.g. the task of arranging photos feels so natural on a multi-touch table is, that it is very close to arranging paper on a real-world table. A user interface that mimics real-world behavior is likely to be intuitive.

Two problems remain with this insight. The first is, that all user interfaces have inherent limitations in their affordances. E.g. 2D touch tables are great in collaborative organisational tasks but poor when 3D manipulation is required, whereas a 3D device like a data glove is a clumsy interface doing annotations. The second problem is, that when users switch between devices to overcome these limitations, they have to interrupt their workflow and adapt to a new method of interaction.

With our system we want to overcome both problems. The first by not trying to build the all-in-one device but by offering a simple, natural way to switch between several devices and domains. The second by making our device do what it does as natural and seamless to the real world as possible by using Augmented Reality (AR) for presentation and bare hand gestures for interaction.

*e-mail: nils.petersen@dfki.de
†e-mail:didier.stricker@dfki.de

IEEE International Symposium on Mixed and Augmented Reality 2009
Science and Technology Proceedings
19 -22 October, Orlando, Florida, USA
978-1-4244-5419-8/09/$25.00 ©2009 IEEE

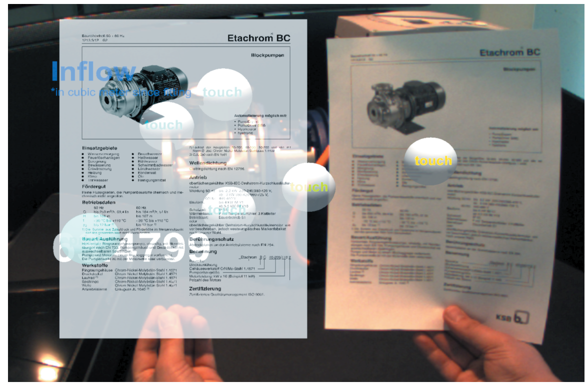

Figure 1: Illustration of domain and interaction continuity between virtual (left) and real-world paper (right) in our demonstrator. Through an Anoto-pattern added during printing all hand-written annotations on the real paper are refllected on the virtual instance.

The remainder of the paper is structured as follows: We will start describing our system followed by a review of related work and connections to other UI concepts. After that, we present the results of the experiments we conducted and eventually give a conclusion of our work.

2 Continuous Natural User Interface

Currently, a user interface is limited to a certain device and data to be processed is loaded withing this interface. We want to reverse this and build a user interface, that naturally spans several devices and keeps centered around the processed information. We distinguish between *content*, denoting the abstract piece of information that the user wants to process and an *instance*, denoting the representation of that within a device or the real world. By spanning several devices, the user can make use of their respective affordances. Furthermore, we not only connect devices but also passive objects that exhibit an embedded intelligence. The most prominent example of such is paper, that offers a natural interface [16], e.g. for receiving handwritten annotations.

To give users the impression, that each presentation is an instance of the same content, all of these instances have to be permanently associated with each other. Additionally, all involved interfaces have to deliver a consistent user experience. Ideally, the user does not feel a breakline when moving from one interface to another. Since we propose a user interface that is continuous between devices, virtual and real-world domain, and in the interaction, we call it Continuous Natural User Interface (CNUI).

2.1 System Description

Our exemplary implementation of a CNUI spans a 3D hand gesture interface, a Multi-Touch interface, and a paper-based instance.

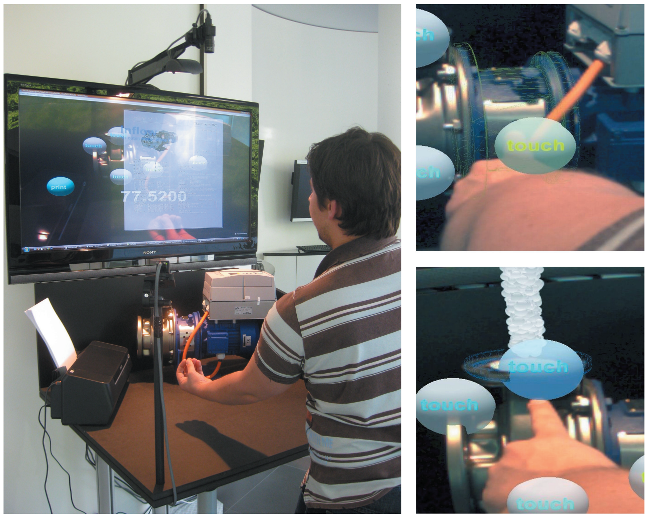

Figure 2: The centerpiece of our demonstrator. The user's hand is tracked in 3D. Through pointing he can select parts of the exhibited object (right). He then can 'grab' information which is visualized as virtual sheet of paper held between the fingers (left).

Centerpiece of our system is the AR-based 3D hand gesture interface for querying and annotating information deposited at real-world objects, see figure 2. The system is built like an info kiosk, exhibiting an object on a table (an industrial pump). Using optically recognized hand gestures, the user can select parts of the object through pointing with the finger. After that he can retrieve information from it through 'pulling' (grab and move away) it out. The information about this part is visualized as a virtual piece of paper held between the fingers, see figure 1. This simulates the behavior of real paper, when held in the hand and also mimics real-world workflow, where information is physically bound and must be taken into the hand to move and place it.

We are using 3D hand gestures that have a real-world counterpart like pointing, grabbing, dragging, and dropping to deal with the virtual paper. Although pointing does not have any physical meaning in the real-world, we are nevertheless using it as a gesture. Firstly, because it is a common way of directing attentiveness to somewhere. Secondly, because it is the most precise way of doing so with the human hand alone. For hand detection we are using [15] with an extension for hand posture classification that is beyond the scope of this paper. The images are captured from two fixed cameras, one above the table, the other on a pole near to the user's head's location which image is displayed on a large screen above the table, see figure 2. This is a compromise to the naturalness of the system, i.e. to not require the user to wear any additional hardware like a head mounted display or camera.

Domain Continuity By dropping the virtual piece of paper on a printer placed next to the exhibited object, the content of the virtual piece of paper is printed together with an Anoto pattern [1]. After printing, the user can uptake the real-world instance from just the same location where he dropped the virtual instance. Additionally, he uses the same grab-and-pull procedure to uptake the real piece of paper, leading to a consistent look-and-feel of the continuated interface, figure 1 illustrates this consistency.

The Anoto pattern enables a corresponding optical pen to identify this piece of paper and precisely locate the pen's position on the paper during hand writings. All written annotations can thus be reassigned to the virtual instance, still deposited at the real-world object. Since the pen is equipped with WiFi, this reassignment is instantaneous. This reinforces the impression of dealing with just another instance of the same content.

With a similarly continuous way real paper can be transferred into the virtual interface: A sheet of paper is being captured by the camera above when put onto the table. It then can be picked up from where it is as virtual instance within the AR-interface and be deposited at a part of the exhibited object. We concentrate on paper based continuity, but the concept would be also possible with other types of objects, e.g. [12]. The cycle of domanin continuity in a CNUI is illustrated in figure 4.

Device Continuity To demonstrate how a continuation between devices could look like, we also connected a touch table standing next to our demonstrator to the system. To transport a virtual sheet from the AR-based interface to the touch table, the user simply moves the sheet out of the demonstrator in direction of the table while still grasped. In this case, the content is inserted into the touch table at the corner, closest to our demonstrator. The user can then access it from there, otherwise it disappears after a short period of time, assuming that the user did not remove it intentionally. This generates the impression that the content was taken along with the user to this device. Nevertheless, since there is no feedback from the interface in the mean time after leaving the one device until entering another, this is only feasible if the two devices are really nearby.

Figure 3: The evolution of user interfaces paradigms and their respective merits. From left to right: Command Line Interface (CLI), Graphical User Interface (GUI), Natural User Interface (NUI), and our approach the Continuous Natural User Interface (CNUI). CLI, GUI, and NUI recreated from [4].

2.2 Concept

We consider the CNUI as being an extension to the Natural User Interface, see figure 3. The possibility of switching scope while upkeeping a degree of continuation for the user helps in hiding the complexity of the system as a whole. The user is less required to abstract from the content, resp. the consistency in the look-and-feel of each instance is supporting the user in making this abstraction. This could enable more efficient workflows and intuitive smart environments.

Especially for a paper based continuation two additional benefits occur. First, the interface is present despite lack of electronical infrastructure and away from the real-world object where the content is associated. Second, it is a permanent hardcopy which can be signed and filed for legal reasons.

To summarize, a Continuous Natural User Interface (CNUI) is:

- **domain continuous**, content transferrable to and from real-world instances
- **device continuous**, content transferrable between all connected devices
- **interaction continuous and consistent**, the way of interaction does not change abruptly when switching between domains and devices
- **content-centered rather than interface-centered**, persistent association between all instances (object oriented user interface)
- **ubiquitous**, the interface can be extended to passive, carriable objects, like paper
- **natural**, interaction is oriented on real-world behavior

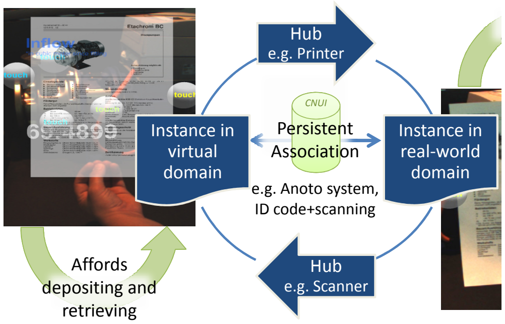

Figure 4: Illustration of a CNUI workflow cycle. This shows, how the user interface is continued between a virtual and real-world domain and makes use of their respective affordances: The two instances stay associated through the use of an Anoto system in this example.

3 Related work

Surprisingly few interaction approaches that focus on natural hand based interaction combined with Augmented Reality have been presented in recent years: [3] combine a 2D multi-touch table with 3D gestures recognized using a data glove and show the 3D part in a head mounted display. They use intuitive natural gestures. Their approach has similarities with the principles we stated about a CNUI. Their focus though, is more on the consolidation of 3D gestures, display, and a touch-table to one user experience, instead of continuation across several isolated interfaces or passive objects. Additionally, they rely on a data glove, whereas our system works without worn hardware. [11] have presented a system for inspection of augmented reality objects. The object thereby is visualized as if it was resting on the palm and the user can rotate and move it by moving the hand. [12] present a system that provides AR overlays on tangible mock-ups made of foam. Their system is marker based and uses skin color cues just to correct occlusions made by the augmentation. [6] use hand tracking and locate Graphical User Interface (GUI) elements on salient parts of real objects in AR to make use of their haptics. Related work also comes from the field of paper augmented work flows: [7] present a system, where they attach fiducial markers to the sides of a sheet of paper and use a ceiling projector to project the screen content on it. Their approach is in aspects similar to ours. While they focus more on the novel possibilities of such lightweight displays, we focus on creating an enveloping user interface. [5] investigated paper augmented workflows, also using an anoto pattern to associate the documents. In the collaborative system of [10], the user deals with an electronic white board with bar code equipped sticky notes. Through scanning those notes, he can access associated content on a computer.

The CNUI is a continuation of a natural user interface (NUI) to other devices or to so-called tangible user interfaces [8]. In the field of NUIs, most work was done towards direction of multi-touch interfaces, e.g. [9], with several systems also available commercially [2, 14]. The multi-touch table in [14] features the possibility of placing objects like a camera on the surface to integrate its content into the interface. Another commercial product worth mentioning is Microsoft's 'Project Natal' [13]. The announcement trailer shows the capturing of a real skateboard to get a virtual look-alike.

4 User Study

We wanted to investigate how quickly users adapt to the new interface and how fast they accomplish a simple task with hand gestures. We compared 1) the performance of our solution to 2) using the mouse with the same type of interaction and 3) the mouse on a standard graphical user interface.

We did a user study with a group of 15 probands, (10 female, 5 male, average age 25.6, age standard deviaton (SD) 10.7) that were recruited from academical resp. non-academical colleagues, visitors of the institute, and acquaintances. None of the probands had any experience with our device. Additionally to the group of first-time users, we had a group of three extensively trained users that did not take part in the questionaire and whose results are listed seperately.

4.1 Experiments and Questionnaire

Each user was asked to accomplish the task of querying and printing information from a real-world object using three different interaction approaches:

- **Our solution (Gesture-Interface)**: Five parts of the exhibited industrial pump (inlet pipe, outlet pipe, flange, motor, and stream-wheel) were selectable indicated through 'touch' labels shown in AR. Probands had to point to one of these, grab the virtual sheet of paper and drop it onto the printer next to the pump to accomplish the task.
- **Mouse controlled (Mouse-Interface)**: Probands were shown the image of the pump with the same 'touch' labels as within the Gesture-Interface and had to pursue the same steps with the mouse. This was to isolate the effect of using the hands compared to using the mouse.
- **Graphical User Interface (GUI)**: Probands had to achieve the task using a mock-up of a standard Windows Vista user interface for the task. This consisted of a file explorer view, showing five PDF-files according to the parts of the pump. After double-clicking on a file a PDF-Viewer opened and users had to click the toolbar's print button. The probands had to announce which part they would choose prior to starting their attempt (and time measurement). So they had to abstract from the real part, to using its name. Since the list only contained five entries the search overhead was negligible.

The steps of the tasks were the same within each interface: 1. pressing a key, to enforce a starting position of the hand and to start time measurement, 2. pointing to a part, 3. grabbing, double-clicking, resp. draging, 4. dropping it onto the printer resp. clicking the print button. The time to accomplish each of these steps was recorded seperately.

The probands were given the scenario that the object is part of an interactive exhibit in a technical museum and the museum was testing three different interaction approaches. This was used to give the probands a concrete scenario, wherein querying of information from real-world objects is useful. Probands were asked to accomplish each task as fast as possible. Though we were interested in performance measurements, by that we actually intended to influence them in perceiving our system more as a comparative interface.

Every experiment was first demonstrated once and users could repeat each task for up to 3 minutes. After the experiments, users were asked to fill out a questionaire. One question covered, whether the proband is using computers extensively in his or her everyday life. After that, probands were asked to score the ease of use, as well as the 'intuitiveness and naturalness' seperately for each of the interfaces on a seven point scale and finally to decide which approach they would prefer as an interface for our given scenario.

4.2 Results

Most users successfully accomplished the task using the gesture interface on the first or second attempt (80 %) and had no problems operating the interface.

The overall results of the performance tests can be seen in figure 5. The six curves show the average times of the best attempts

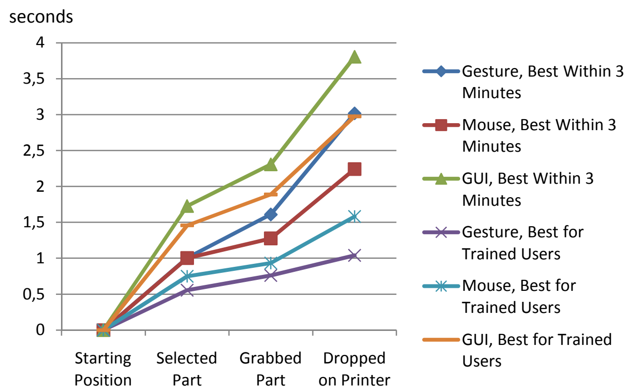

Figure 5: Best attempts within 3 minutes for first-time users and trained users for all investigated interfaces. The best performance was achieved by trained users using our gesture interface.

for each experiment, both of first-time and trained users. In both groups, the task took longer using the graphical user interface (GUI) than with the mouse on the pump's image, due its larger targets. After three minutes, first-time users still were fastest using the mouse on the pump's image, while all probands from the trained users group were significantly faster using the gesture interface.

The question for ease of use of the three approaches did not lead to any clear favorite. Since the subjective ease of use of a longly trained interface, like mouse and GUI and of an unknown, new approach can only be superficially investigated during a short-time study, this should be a focus of a future long-term study. Also the task itself was very simple to achieve using all three interfaces.

For the question about the 'intuitiveness and naturalness' of the three approaches, surprisingly many users chose the GUI as being very intuitive: half of the probands gave a score of 6 or the maxmimum 7. Interestingly, we found a correlation between the self assessed computer knowledge score with the scores given for 'intuitiveness and naturalness' of the GUI and for mouse interaction on the pump's image. This correlation was evident in given statements and was backed by statistics (GUI: Correlation coefficient $\rho_{GUI} = 0.33$, on a significance level of $\alpha_{GUI} = 0.02$, mouse interaction on pump's image: $\rho_{mouse} = 0.43$, $\alpha_{mouse} = 0.02$). To explain this: Most of our probands were very used to graphical user interface (8 probands gave themselves the maximum score 7 in the self-assessment of computer knowledge, avg: 5.8 sd: 1.6, only). One proband even complained about, our GUI mock-up not supporting his commonly used keyboard shortcut sequence for printing, still giving a high score for the 'naturalness' of the GUI approach. Obviously, users are so used to graphical user interfaces, that they perceive the incorporated metaphors and workflows (including keyboard shortcuts) as being the natural way for processing information.

On the question, which user interface they would prefer, 80% (12 probands) gave the gesture interface as answer. The remaining 20% (3 users) decided for the mouse interface on the pump's image (all with the maximum computer knowledge score of 7). This is surprising, since all first-time users were faster using the mouse. Still most enjoyed the gesture interface considerably more and felt this to be more 'direct' for dealing with real-world objects than an abstraction through a graphical representation.

5 Conclusion and Future work

A novel approach to a natural user interface that connects devices and plain objects like paper to a continuous user experience has been presented. Centerpiece of our system is a 3D-gesture driven interface for querying information from real-world objects, using Augmented Reality techniques for presentation. We demonstrated our approach on a workflow using this 3D-gesture interface, a touch table, and a paper based continuation of the interface. In a user study with 15 probands we investigated the acceptance of hand gestures interfaces compared to mouse based ones.

Interestingly, the subjective feel of 'Intuitiveness and Naturalness' of the two mouse based approaches strongly correlated with the self assessed computer knowledge score. 80% of the probands prefered the gesture interface over the other provided, mouse based approaches and 100% with a computer knowledge score lower than the maximum of 7.

Future work includes investigating continuous workflows in an actual work evironment, like the industry. Also, since we expect some results being biased by the short-time nature of our study, we propose a long-term study for that. On the interface side, we would like to extend our prototypical implementation to incorporate more distinguishable gestures in our system and to comprise non-paper real-world instances.

Acknowledgements

This work was supported in part by a grant from the German BMBF (project AVILUSplus, grant number 01IM08002).

References

[1] Anoto Group. Anoto digital pen. http://www.anoto.com.
[2] Apple. iPhone Multi-touch. http://www.apple.com/iphone.
[3] H. Benko, E. Ishak, and S. Feiner. Cross-dimensional gestural interaction techniques for hybrid immersive environments. In *Proc. VR '05*, pages 209–216. IEEE, March 2005.
[4] A. de los Reyes. Predicting the past. http://www.webdirections.org/resources/august-de-los-reyes-predicting-the-past.
[5] F. Guimbretière. Paper augmented digital documents. In *Proc. UIST '03*, pages 51–60, New York, NY, USA, 2003. ACM.
[6] S. J. Henderson and S. Feiner. Opportunistic controls: leveraging natural affordances as tangible user interfaces for augmented reality. In *Proc. VRST '08*, pages 211–218, New York, NY, USA, 2008. ACM.
[7] D. Holman, R. Vertegaal, M. Altosaar, N. Troje, and D. Johns. Paper windows: interaction techniques for digital paper. In *Proc. CHI '05*, pages 591–599, New York, NY, USA, 2005. ACM.
[8] H. Ishii and B. Ullmer. Tangible bits: towards seamless interfaces between people, bits and atoms. In *Proc. CHI '97*, pages 234–241, New York, NY, USA, 1997. ACM.
[9] S. Izadi, S. Hodges, S. Taylor, D. Rosenfeld, N. Villar, A. Butler, and J. Westhues. Going beyond the display: a surface technology with an electronically switchable diffuser. In *Proc. UIST '08*, pages 269–278, New York, NY, USA, 2008. ACM.
[10] B. M. Lange, M. A. Jones, and J. L. Meyers. Insight lab: an immersive team environment linking paper, displays, and data. In *Proc. CHI '98*, pages 550–557, New York, NY, USA, 1998. ACM Press/Addison-Wesley Publishing Co.
[11] T. Lee and T. Hollerer. Handy ar: Markerless inspection of augmented reality objects using fingertip tracking. In *Proc. Int. Symp. on Wearable Computers*, pages 83–90, 2007.
[12] W. Lee and J. Park. Augmented foam: a tangible augmented reality for product design. In *Proc. ISMAR '05*, pages 106–109, Oct. 5–8, 2005.
[13] Microsoft. Project Natal. http://www.xbox.com/projectnatal.
[14] Microsoft. Surface. http://www.surface.com.
[15] N. Petersen and D. Stricker. Fast hand detection using posture invariant constraints. In *Proc. KI 2009: Advances in Artificial Intelligence*. Springer, 2009.
[16] A. J. Sellen and R. Harper. *The myth of the paperless office*. Cambridge, Mass. : MIT Press, 2001.