NUI-Based Floor Navigation — A Case Study

Ulrich Furbach and Markus Maron

Department of Computer Science, Artificial Intelligence Research Group
University of Koblenz-Landau, Universitätsstr. 1, 56070 Koblenz
{uli,maron}@uni-koblenz.de

Abstract. In this paper, we describe a nui-based application using a Microsoft Kinect. The system displays a digital representation of a university building, where users can navigate virtually through contact-less gestures. Users can step up and couple their hand with a virtual mouse cursor to navigate through the program such that hand movements to the right lead to cursor movements to the right for example. We present an evaluation of the system, which is based on a 100' day operation by logging 2.000 user sessions.

1 Introduction

Natural user interfaces is a well researched topic over the past years. Gestures play a central role for contact and non-contact interfaces as well. In particular, the huge success of smart phones fosters a lot of innovative development.

However, there is also the need for contact-less gestures, e.g. within an operating room or behind a shop front window[11]. The number of smart ideas and applications of contact-less gestures have exploded since the availability of the Microsoft Kinect[6]; a piece of hardware which is cheap, and the OpenNI Framework[10] easy to programand easy to embed into complex systems.
We describe the design and evaluation of a system for NUI-based floor navigation within a campus building. Users can navigate virtually through contactless gestures. They can step up and couple their hand with a virtual mouse cursor to navigate through the program such that hand movements to the right lead to cursor movements to the right for example. The system is part of a campus-wide information system, which is used to display various kind of information[1].

Located in the entrance area of a newly constructed building users unfamiliar with this should find their way around quickly. Therefore, the goal was to develop an innovative interactive application, which empowers their users to acquire detailed information on floor levels and individual rooms, such as names of employees and contact data. Fulfilling the HCI design constraint "Come as you are" [14] , the application allows users to navigate through the building just with the help of gestures without any additional requirements needed, moving a virtual hand over a floor map displayed on an up-right widescreen TV mounted in a brushed iron frame fixed to a wall (cf. Figure 1). Using gestures instead of

[1] http://www.wizai.com/index.php/loesungen/campusnews

A. Marcus (Ed.): DUXU/HCII 2013, Part III, LNCS 8014, pp. 270–279, 2013.

Fig. 1. Wall-mounted gesture control floor navigation system

a touchscreen makes the content on the screen accessible for everyone, also for small or handicapped people.

In the context of public displays to motivate and focusing the attention [3][4][5][8] of the passers-by to the display are well discussed topics. In our case there are three main challenges:

- *Users do not play a game.* The user who stops in front of the floor plan wants to solve a specific task, namely looking for a room or a person. They are generally not willing to spend time to learn or to experience something new.
- There is *no chance to teach* the user or to make them read a manual before using the device.
- There are *no commonly accepted gestures* for controlling a screen - we have to assume no prior user experience.

Specially the last point turned out to offer a real challenge for the design of the gesture interface. Many users find themselves rather helpless with regard to the system. During the design phase of the system we did some experiments in order to find the most appropriate gestures. After installing the system, we collected data about the usage in log files along with videos of the users's behavior. We will offer an evaluation of the first few months of the application in a public building on campus.

2 Related Work

The scene in Steven Spielberg's *Minority Report* science fiction movie is well-known where Tom Cruise uses gesture control to manipulate images. This was unimaginable in 2002, but now 10 years later it is a reality. Samsung has just released a new Smart TV with voice and gesture control. In the field of natural user interfaces a lot of research has been conducted not least since Microsoft released the Kinect: A cheap and robust sensor and a SDK for developing. More than 100.000 individuals downloaded the SDK in the first six weeks. Using a display with gesture control instead of a touchscreen offers the chance to install an interactive display behind a shop front window for presenting their goods or just analyzing the user behavior.[11] Besides doing research and using it within home entertainment, the usage of the Kinect can be useful in several scenarios where input with controller or touch are not useful. In the medical field it is used to manipulate medical images without having to touch a controller[1], reducing the chance of hand contamination in operating theatres[7][15]. For using it in such a critical environment it is important that the handling is as simple as possible. But finding simple and intuitive gestures is not trivial. "*Poke it or press it*, everybody had a very different idea of what that actually meant." [2]

3 The Application

The system is located in the entrance area of a newly constructed university building. Users unfamiliar with the new building should find their way around quickly. Therefore, the goal was to develop an innovative interactive application, which empowers their users to acquire detailed information on floor levels and individual rooms, such as names of employees and contact data. The application allows users to navigate through the building with the help of gestures, moving a virtual hand over a floor map displayed on an upright widescreen TV fixed to a wall. Using gestures instead of a touchscreen enables the usage of the entire display for everyone, also for small or handicapped people.

The floor plan of the entire building, including all rooms, was to be coherently displayed in an application and run permanently on a mini linux computer. The navigation through rooms and floors is enabled by gesture controls. The following gestures were to be implemented: Wave, push and swipe. Each of these are then associated with actions to enable navigation as shown in the following table.

In order to allow the selection of certain rooms, the user's hand should be coupled to a virtual mouse cursor on the screen so that objects can be selected

Table 1. Defined gestures and their calling actions. *Push is realized by holding the hand 4 sec. above the clickable element. (cf. Figure 3)

gesture	action
Wave	Activate
Swipe_Left, Swipe_Right	Switch_Person
Swipe_Up	Switch_Floor_UP
Swipe_Down	Switch_Floor_Down
Push*	Entered_Room, Left_Room, Floorbox_Pushed, HelpButton_Pushed

on the screen similarly to the way objects are selected with a normal mouse on a computer. An object should be selected by an appropriate gesture. Depending on the selected object, different information can be displayed. For lecture halls, this information contains the name of the current lecture being held, the person holding the lecture, and the subsequent lecture. For offices, this information includes the employee's name, his/her contact information, an avatar or photograph, and a QR-Code with condensed information of that person. The necessary contact data for all employees and lectures can be updated every night and saved in a database.

Implementation. The system development was separated into two parts. The floor plan application and the development of the gesture recognition and control of the application. The entire program is written in C++ with the help of the OpenFrameworks toolkit.

All necessary employee and lecture information for the floor plan part is retrieved from a database so that only up-to-date information is displayed.

For the gesture recognition, the SensorKinect driver by Primesense[12] was used in combination with OpenNI (Open Natural Interaction)[10], a framework which provides several different APIs for natural interaction devices. Additionally, NITE (Natural InTEration) was used. This framework also provides APIs for interaction between humans and machines. By combining these three technologies it is possible to read and analyze Kinect data. OpenNI provides functionality so that new data from the Kinect can be analyzed and gestures identified.

Fig. 2. Video help for session activation. Translated into English: "Hand–control: Wave. Please two meters distance."

Gesture design. As mentioned in the introduction, there are three main characteristics of this application. During the design of a prototype we had to address all three of them:

– *Users do not play a game.* A user stops in front of the screen in order to get information quickly. At this moment they do not know that the screen can be controlled by gestures. This is in the field of public displays a very common problem[9]. In order to clarify this, we run during the standby phase permantely a movie in the lower part of the screen (cf. Figure 1), which shows a hand waving permanently together with the written info, that this is the way to activate mouse-control manually (cf. Figure 2). After the user's hand is recognized and tracked, the user can move the curser. This turns out to work nicely, however many users put down their hand after activation instead of controlling the curser. They simply expect another action from the system.

Fig. 3. Mousehoover feedback for clicking actions

- *No chance to teach the user.* In a first approach during the development of the system we offered our test users wiping and pushing gestures. However, it turned out to be rather difficult to offer these gestures. We learned that users do not read any further help which is displayed on screen. Moreover, they immediately try to use individual gestures. Recognizing and scaling these gestures appears to be too difficult for a practical application. Therefore, we decided to use a rather traditional approach, clicking is implemented by mouse-hovering, depicted in Figure 3.
- *No commonly accepted gestures.* During the experimental phase of the system design we learned that an average user has a lot of problems in using gestures for navigation if there is no instruction. We will discuss this point in the following evaluation.

4 Evaluation

In this section we describe the evalution based on collected data within 145 days about the usage in log files along with 2,245 videos of the users's behavior. These data gave us a insight in the behaviour and the emotions of the user interacting with a NUI-based control. The video analysis brought out that man and female users are dealing in different ways with that nui-based floor control. Besides some interesting gender aspects we observed group dynamics as well.

4.1 Usage of the System

Figure 4 shows the distribution of the usage during the last months. The number of sessions and actions decreased during the semester and reached their minimum in the semester break. Afterwards, both figures started to increase again. The number of actions per session is nearly constant about 6 (see Figure 5).

This development during the last 5 months proves the acceptance of the system as a daily routine. We will investigate this in more detail in the following.

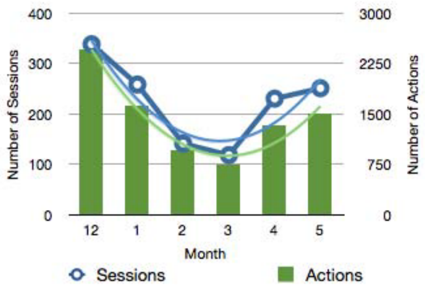

Fig. 4. Usage during the time period: shows the number of sessions and the number of actions per month

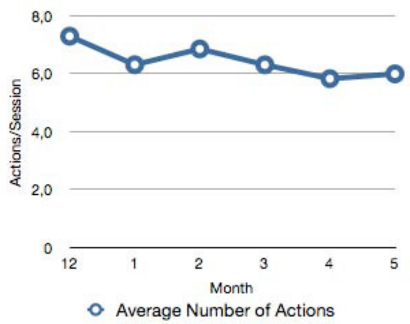

Fig. 5. Average number of actions per session during the time period

4.2 Sessions, Actions and Events

Overall, since the rollout till now (14th of May) a total of 2.065 sessions have been started. On closer inspection, we have detected that 165 sessions have been opened unintentionally by people standing close to the system while talking to another person gesturing with their hands. Additional, 368 sessions have been recorded which contain no opening action. This means that during these sessions the person in front of the system tried to activate the control but did not succeed. Additional, 469 sessions have been successfully opened but the interacting person did not recognize the announcement on the display. If we adjust the logs and reduce these failure sessions we count 1.202 successfully opened sessions with 7.833 actions and 4.751 events. Unfortunately, we count 3.083 actions which did not lead to an event.

Figure 6 depicts the distribution of actions. The most performed action was *Floorbox_Pushed* and, interestingly, the HelpButton was pushed only 19 times, even though it is placed very visibly.

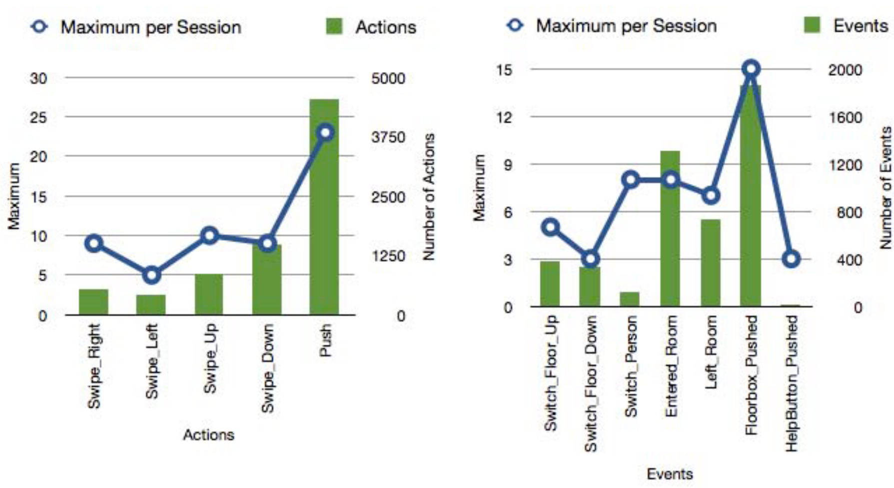

Fig. 6. Total number of actions (left) and events (right) and the maximum number of each action/event performed per session

Figure 7 presents the distribution of the number of actions per session. 60% of all people performed more than 5 actions. At maximum, one person performed 34 interactions within 79 seconds another person spent 181 seconds while doing 18 interaction steps. In total, all users spent 36.380 seconds accordingly 10h 6min. The average usage time is 30.2 seconds per session and 4.6 seconds per action.

Figure 8 shows that most of the users spent more than 10 seconds within a session. This is not because of the unfamiliarity with the user interface, as shown in the following evaluation of the recorded videos.

4.3 Observing the Users

The videos, we recorded for a more careful semantical evaluation, show that many users performed exaggerated motions in front of the system at the beginning of a session, but after a short while they learned how to control the system.

Indeed, learning-by-doing is the most important factor in the shift from novice to experienced user.

We analyzed the videos of 188 sessions during a period of 17 days. In these sessions we counted 176 different people standing in front of our camera, 88 people interacted with the application (Figure 9). In 129 sessions the interacting person was accompanied by other people. The maximum was a group of 5.

The video analysis also showed some gender aspects. 33 of the interacting people were female and 55 male. Males performed more actions and harder than

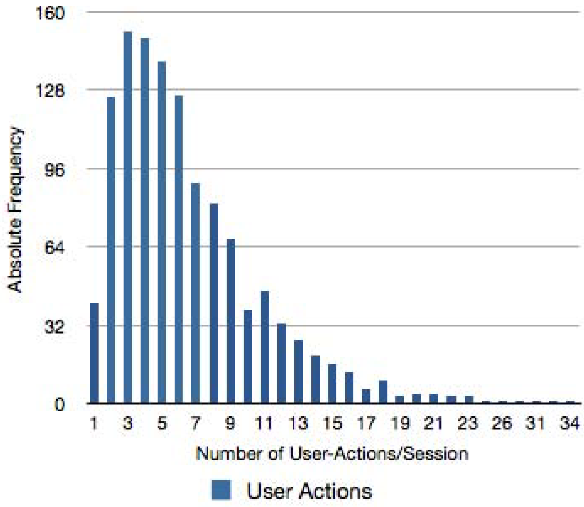

Fig. 7. Distribution of number of actions per session

females. The maximum amount of actions was 16 performed by a male (11 female) and the average amount of actions per session was 7.7 by males (5.0 female).

Most of the interacting persons showed positive emotion. 85 percent of the females and 78 percent of the males left the place with a smile on their face.

5 Lessons Learnt

The development and the evaluation of the system reported in this paper started as a student project. In the beginning a lot of experiments have been done in order to find easy and precise gestures for the specific task of a floor navigation system. During this initial phase it turned out that this is by far not trivial. For example, we thought that waving is a good and simple gesture to activate the application control. But we had to learn that people wave hands in their own way and a lot of them did not achieve to take control of the application. From this experience we came to the solution to show the activating waving-gesture in an introduction video, which is shown whenever the screen is not in use.

When we finally mounted the system on a wall in the entry area of the building, we learnt a lot about changing lighting in the building in the course of an entire day and about its influence on the performance of the system. Also, the area in which the Kinect should identify users and react to their actions has to be determined by numerous experiments. Then we started the evaluation phase in which we collected the data which was evaluated in the previous section.

The main points from this evaluation are

– Since there is no chance for such a system to train users, it is important that learning can be done during a single session. The number of actions necessary

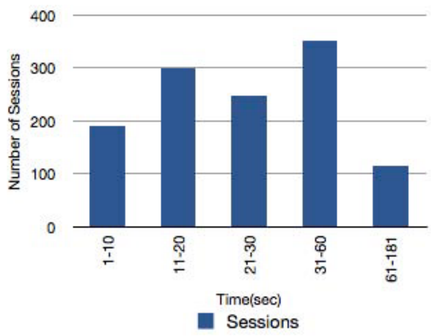

Fig. 8. Duration time user spent

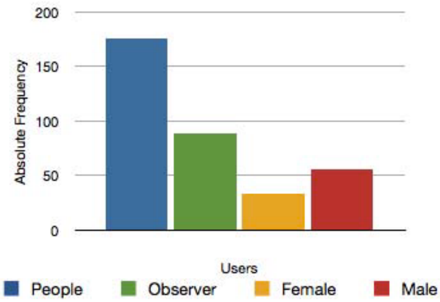

Fig. 9. Distribution of interacting people

to perform an event is usually decreasing during a single session, which clearly indicates that the user learnt to control the system more efficiently.
– Our evaluation during several months proves that such NUI-based systems are ready to be used in real-life applications under realistic and natural conditions.
– The video analysis of a smaller sample gave us additional insight into the behavior of users. Although this analysis is of course rather limited, because it is based on interpretations of the assessor, it can be used very well as a kind of formative empirical evaluation.

For us it was fun to develop the application and for most of the people using it, it is was fun too [13]. A more detailed description and evaluation we will give in an other paper.

Acknowledgments. We thank the University of Koblenz for giving us the opportunity to install the system and thus having the chance to use thousands of users for this research.

References

1. Gallo, L., Placitelli, A.P., Ciampi, M.: Controller-free exploration of medical image data: Experiencing the Kinect. In: CBMS, pp. 1–6. IEEE (2011)
2. Goth, G.: Brave NUI world. Commun. ACM 54, 14–16 (2011), http://doi.acm.org/10.1145/2043174.2043181
3. Huang, E.M., Koster, A., Borchers, J.: Overcoming assumptions and uncovering practices: When does the public really look at public displays? In: Indulska, J., Patterson, D.J., Rodden, T., Ott, M. (eds.) PERVASIVE 2008. LNCS, vol. 5013, pp. 228–243. Springer, Heidelberg (2008), http://dx.doi.org/10.1007/978-3-540-79576-6_14
4. Jacucci, G., Morrison, A., Richard, G.T., Kleimola, J., Peltonen, P., Parisi, L., Laitinen, T.: Worlds of information: designing for engagement at a public multi-touch display. In: Mynatt, E.D., Schoner, D., Fitzpatrick, G., Hudson, S.E., Edwards, W.K., Rodden, T. (eds.) CHI, pp. 2267–2276. ACM (2010)
5. Michelis, D.: Interaktive Großbildschirme im öffentlichen Raum: Nutzungsmotive und Gestaltungsregeln. Gabler Edition Wissenschaft, Gabler Verlag (2009), http://books.google.de/books?id=Q9UkyR--HE0C
6. Microsoft Kinect, www.xbox.com/en-US/kinect (last accessed on February 26, 2013)
7. Moretti, S.: Doctors use Xbox Kinect in cancer surgery (March 2011), http://tinyurl.com/4ym6866 (last accessed on February 26, 2013)
8. Müller, J., Walter, R., Bailly, G., Nischt, M., Alt, F.: Looking glass: a field study on noticing interactivity of a shop window. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2012, pp. 297–306. ACM, New York (2012), http://doi.acm.org/10.1145/2207676.2207718
9. Müller, J., Walter, R., Bailly, G., Nischt, M., Alt, F.: Looking glass: a field study on noticing interactivity of a shop window. In: Konstan, J.A., Chi, E.H., Höök, K. (eds.) CHI, pp. 297–306. ACM (2012)
10. OpenNI - The standard framwork for 3D sensing, www.openni.org (last accessed on February 26, 2013)
11. Peltonen, P., Kurvinen, E., Salovaara, A., Jacucci, G., Ilmonen, T., Evans, J., Oulasvirta, A., Saarikko, P.: It's mine, don't touch!: interactions at a large multi-touch display in a city centre. In: Proceedings of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems, CHI 2008, pp. 1285–1294. ACM, New York (2008), http://doi.acm.org/10.1145/1357054.1357255
12. PrimeSensor™ Reference Design, http://www.souvr.com/Soft/UploadSoft/201005/2010050617295050.pdf (last accessed on February 26, 2013)
13. Shneiderman, B.: Designing for fun: how can we design user interfaces to be more fun? Interactions 11(5), 48–50 (2004)
14. Triesch, J., von der Malsburg, C.: Robotic gesture recognition. In: Wachsmuth, I., Fröhlich, M. (eds.) GW 1997. LNCS (LNAI), vol. 1371, pp. 233–244. Springer, Heidelberg (1998)
15. Wall, J.: Kinect and Medical Imaging, www.youtube.com/watch?v=U67ESHV8f_4 (last accessed on February 26, 2013)