Easy collaboration on interactive wall-size displays in a user distinction environment

Marc Turnwald  
Information and Technology Management  
Ruhr-University of Bochum  
Universitätstraße 150  
44780 Bochum  
+49 234 32 22045  
turnwald@iaw.rub.de

Alexander Nolte  
Information and Technology Management  
Ruhr-University of Bochum  
Universitätstraße 150  
44780 Bochum  
+49 234 32 22045  
nolte@iaw.rub.de

Michael Ksoll  
Information and Technology Management  
Ruhr-University of Bochum  
Universitätstraße 150  
44780 Bochum  
+49 234 32 4442  
ksoll@iaw.rub.de

ABSTRACT
This paper describes the prototype of an attachment-free user distinction system for multi-user interaction with interactive wall-sized displays. The system offers support for situations in which easy and synchronous collaboration of multiple users is required. We based the development upon one typical scenario for synchronous collaboration, which is the clustering of previously collected brainstorming items with a group of participants. We developed and tested a prototype for this scenario, present preliminary results and describe future requirements for development as well as directions for research.

Categories and Subject Descriptors
H.1.2 [User/Machine Systems]: Human Factors  
H.5.2 [User Interfaces]: Input devices and strategies, User-centered design

General Terms
Design, Human Factors

Keywords
Collaboration, Participation, Multi-User Distinction, Kinect

1. INTRODUCTION
During the last few years research upon wall-sized interactive displays has gained a lot of interest. A number of applications emerge aiming at exploiting the unique advantages of these displays. One very important stream is to support groups of multiple users to work on the same artifact at the same time. This approach promises to create more involvement among workshop participants as it allows for a direct interaction with the displayed materials by all of them [3]. This in turn often bears the necessity to recognize and distinguish multiple interactions by multiple workshop participants at the same time.

In order to achieve this without the necessity of identification or through individual attachments using e.g. mobile devices [13], we used the Kinect system by Microsoft. This system is capable of distinguishing and tracking of up to 7 people at the same time. Combined with an interactive large screen and a corresponding client-server architecture this results in a low-cost attachment-free multi-user distinction system.

The scenario we are using for this multi-user distinction system is the design of processes through collaborative process modeling [11]. This includes designing processes from the scratch as well as re-designing existing ones. When processes or process parts have to be designed from the scratch, it is useful to include means of creativity, e.g. brainstorming [6]. This, in turn, means that a large number of contributions has to be clustered, which can be very time-consuming when done by a single person. So we developed the idea to let multiple people do the clustering at the same time.

Collaborative clustering is not the only purpose within which a multi-user distinction system is useful in the context of collaborative modeling on wall-sized displays. It would also provide a lot more flexibility when it comes to assessing the resulting model by using a walkthrough oriented approach [5]. Multiple sub-groups could work on different parts of a model simultaneously within the same representation, thus speeding up the modeling process and allowing people to focus on the parts of the model they are really interested in, rather than having to wait for their part of the model to become the focus of a conversation [9].

In the following, we provide a short literature review (section 2). Afterwards, we derive requirements based upon the aforementioned scenario, describe the environment in which the system was developed and used, followed by a description of the prototypical setting (section 3). We conclude with an outlook upon future development and research (section 4).

2. RELATED WORK
2.1 Collaboration on wall-sized displays
Conceptually related work can be traced back to the first appearance of the term Single Display Groupware (SDG) [14] in 1999. Stewart observed multiple users cooperating in a co-located setting with a single display through providing an individual input channel for every user (namely a mouse and a keyboard). Since the emergence of wall-sized interactive displays a lot of research has been done considering multiple users working at the same display at the same time [8]. While technical issues with the first generation of touch detection systems have been solved in the past, there are still only a few systems exploiting their capabilities including systems for collaborative voting [4].

2.2 Distinction mechanisms
The first touch sensitive display with the ability to distinguish its users was the MERL DiamondTouch [1]. It consists of a tabletop display that was able to detect multiple users and to distinguish them by their capacitive resistances. More up to date approaches extend touch sensitive surfaces with vision based sensory. Dohse et al. [2] used a camera based system that was placed above an interactive table to detect the hands of users and thereby their position around it. An even more sophisticated method was described by Schmidt et al. [12], within which user identification is done by identifying biometric differences between the different hand
shapes of the user’s interaction with it. Research on these systems however was limited to tabletop displays. Vertical displays, such as the one we use (c.f. Figure 1), have not been examined yet.

For vertical displays Schöning et al. [13] use mobile phones to identify users. Other solutions use different attachments to the hands of the users, thus detecting these attachments to distinguish multiple users interacting with the surface. Rekimoto [10] used digital pens to provide user IDs related to touches. This, however, bears the necessity of users wearing additional devices or attachments, which we wanted to avoid within our setting.

Newer prominent research on the combination of interactive displays and the kinec is described by Wilson and Benko [15]. A system, that is very similar to ours, was recently proposed by Jung et al. [7]. Beside similarities to our system, their system is built around a horizontal display and proposes also additional RFIDs for user authentication. The necessity of authentication is also the main difference to our approach, namely to provide a system for easy participation in which users can be distinct, but need not to authenticate.

Besides using additional systems for user identification, users could also identify themselves. They could e.g. define an area that they want to work in. The result of such a scenario is a system, which is capable of distinguishing between each touch that occurs by the user’s working area, which she has set up before. This approach, however, is not suitable for our scenario, as it limits user interaction to certain parts of the screen. After each interaction, the users could also identify themselves through e.g. selecting their identifier out of a list of users of the system [4]. This requires additional touches after each interaction, thus significantly slowing down the interaction with the system.

3. PROTOTYPE

3.1 Requirements

In order to match the aforementioned clustering scenario within the context of co-located modeling workshops, the system has to meet the following requirements:

* Distinguish multiple users: Clustering requires people to pick up cards and place them somewhere else. As multiple people will perform this task at the same time, the system has to be able to distinguish between them at any given time.

* No additional attachments: Using personal attachments or mobile devices for user distinction leads to an extended preparation time. However, as clustering might be useful at any time during a process development effort, it is desired to reduce preparation time as much as possible in order to provide a smooth transition between workshop phases. So, the system has to be able to distinguish multiple users without additional attachments.

* No need for identification: Using a system similar to [4] requires the user to identify herself after each interaction. This would significantly slow down the clustering process. Thus, the system has to be able to distinguish multiple users without the need for identification.

* No space limitation: The clustering scenario requires multiple people to work on the same artifact at the same time. So, the system has to allow all these people to work over the whole width of the wall at any given time, thus making identification through a working area as described in section 2.2 impossible.

It has to be noted that these requirements are not strictly limited to clustering, but could also apply to a number of other collaboration tasks on interactive large screens such as e.g. voting [4].

3.2 Setting and hardware

The prototype we developed was created to be tested in a special facilitation co-laboratory, called ModLab (c.f. Figure 1). Its centerpiece is a large, high-resolution interactive screen (4.80m x 1.20m; 4200x1050px) consisting of three seamlessly connected rear projection boards. The whole surface is touch-sensitive and thus allows seamless operations over the whole width of the wall.

Figure 1: The ModLab – University of Bochum

![Figure 1: The ModLab – University of Bochum](image)

Figure 2: Client-Server Architecture for multi-user distinction on interactive large screens through a Kinect system

In addition to the interactive large screen, we use a Microsoft Kinect sensor to track users interacting in front of it. The Kinect is capable of detecting, distinguishing and tracking human bodies through a time-of-flight camera that generates a 3D point map the scene it is directed at. Each point of such a point map can be complemented with a user ID, provided that it belongs to a detected human body form. In order to observe the whole width of the
wall, we placed the Kinect about 3 meters away from the interactive wall right below the ceiling of the room (c.f. Figure 2 center).

3.3 Realization of the software

The implemented prototype is based on the Processing programming language framework. It consists of two different applications that are connected with each other via LAN. The first application, which runs on the computer that powers the large screen as well, is a simple clustering application. The second application, which runs on an additional PC, is connected to the Kinect and thus is responsible for distinguishing and tracking the user’s interaction with the large screen. For the Kinect’s operation purposes we used the OpenNi programming framework via the SimpleOpenNi wrapper for the Processing programming language framework.

Within our setting, a user’s body is recognized by the Kinect and virtually represented as an array of 3D points (c.f. Figure 3), when she enters the scene in front of the wall. When she touches the screen in order to pick a card, this touch is detected by the clustering application and its 2D coordinates are send to the user distinction sensory server (c.f. step 1 in Figure 2).

![Figure 3: 3D representations of four users standing in front of the large screen](image)

Afterwards, these coordinates are transformed into 3D coordinates (c.f. step 3 in Figure 2), thus representing the touch position in 3D. Prior to this, the system has to be calibrated, in order to execute this transformation. We intentionally miss out on describing the calibration as it becomes obsolete once the Kinect is mounted at a fixed position. Hereafter, the shortest distance between each user (represented by an array of 3D points) and these coordinates is calculated (c.f. step 3 in Figure 2). Having detected the nearest user, the original 2D coordinates are enriched with a corresponding user ID and sent back to the clustering application running on the large screen (c.f. step 4 in Figure 2). When the clustering application receives a position, which is enriched with a user ID, the corresponding event is processed (c.f. step 5 in Figure 2).

The event processing is done according to the state of the corresponding user’s pick-up slot and the hit position, where the user touches the large screen (c.f. Table 1). The pick-up slot is a virtual representation of her hand and thus references a picked up item, like a card or a cluster. It can be either empty or occupied. The hit position can also either be empty or occupied by an item like a cluster or card. Table 1 shows the processing results for each hit position and pick-up slot state. So e.g., when a user starts working with the system (c.f. Table 1: pick up slot / empty) and wants to place a card into an already existing cluster, she has to select it, thus “picking it up” (c.f. Table 1: hit area / occupied by card). Now, as she holds the card in her hand (c.f. Table 1: pick up slot / occupied by card) she has to tap the cluster where she wants to place the card (c.f. Table 1: hit area / occupied by cluster). This results in the card being moved from its original position into the selected cluster. Table 1 shows the processing results for each hit position and pick-up slot state.

<table>
<thead>
<tr>
<th>hit area</th>
<th>state of user’s pick up slot</th>
</tr>
</thead>
<tbody>
<tr>
<td>empty</td>
<td>occupied by card</td>
</tr>
<tr>
<td>occupied by cluster</td>
<td>drop card</td>
</tr>
<tr>
<td>occupied by card</td>
<td>drop cluster (anyway)</td>
</tr>
<tr>
<td>occupied by tool button</td>
<td>press button</td>
</tr>
</tbody>
</table>

Table 1: System reaction according to pick up slot and hit area.

Based upon the previously described requirements, the application is only capable of handling simple touch events. Apart from allowing the application to run smoothly, this also results in running it even on systems that are only capable of detecting a single touch event at the same time.

3.4 Test of the prototype

For calibration purposes we have to pin Post-it notes to the corners of the interactive surface (c.f. Figure 4). These Post-it notes are necessary to improve the detection of the borders of the interactive surface by the Kinect sensor. Afterwards, the 3D coordinates of these corners have to be marked in the displayed scene (c.f. Figure 3) with a cursor. This calibration is required for the transformation from 2D into 3D coordinates to in order to work properly (c.f. step 2 in Figure 2). Hereafter, the post-its can be removed. Once a calibration is done, this step gets obsolete, provided that the Kinect is left at a fixed position in front of the large screen.

![Figure 4: calibration marker in the top left corner of the interface](image)

Before the users started working, we described the system’s functionality by telling the users how to create clusters, how to move them and how to place cards within them. We put special emphasis into explaining that only tap operations were allowed and that drag operations would result in the system to intermit. Afterwards, we advised them to start using the system. During the test, which took about 45 minutes in total, we switched between different constellations of participants. We also altered the number of users interacting with the system at the same time, ranging between two
and four users. After the test, we conducted a short group interview disclosing what the participants liked about the system, e.g. intuitive usability, goal-oriented multi-user collaboration or fostering of idea creation and clustering. They also provided ideas on how to improve the system like the combination of multiple Kinect devices and intensified use of colors for distinguishing purposes of user activities.

3.5 Preliminary results

The test revealed that the prototype was working stable and fulfilled its main tasks. The system was able to differentiate between users and to distinguish between their interactions within the clustering application. This meant, for example, that one user was able to create a cluster while another user was moving cards around simultaneously (see Figure 5). Moreover, even four users were able to use the system in the same way.

However, we also observed that an increasing number of participants resulted in the prototype becoming inaccurate. While two users could operate the system smoothly, four users often caused the system to mix up different participants, thus losing the connection between the touch interaction and the corresponding user ID. This resulted in participants moving the items of other around, hence leading to confusion.

Figure 5: Three users operating the clustering tool in parallel.

These infractions came due to our setup consisting of one Kinect that is placed far away from the large screen. So, when users stood close to each other, the system often failed to distinguish between them. This was especially the case, when two users, who differed significantly in size, operated close to each other.

4. CONCLUSION AND OUTLOOK

After having successfully developed a system that couples an interactive wall-sized display with a Kinect system for user distinction, we have tested it in a clustering scenario. This test showed a number of difficulties, highlighting the issue of overlapping people as the most serious one. In order to overcome this problem, we are planning to set up multiple Kinect systems watching the interactive large screen from different angles. Syncing the corresponding 3D representations should lead to an increased distinction mechanism.

Apart from future development, we are also planning on a proper experimental setting in order to measure the benefits of this system compared with the traditional clustering approach, within which a facilitator does the clustering based upon comments by the audience. We are also planning to use this distinction system for other activities within the context of collaborative modeling, such as parallel development of multiple model parts by small groups directly at the large screen.

5. REFERENCES