Deviceless Gestural Interaction in Public Displays

Thiago Motta, Luciana Nedel
Institute of Informatics – UFRGS, Porto Alegre, Brazil
Email: {tsmotta,nedel}@inf.ufrgs.br

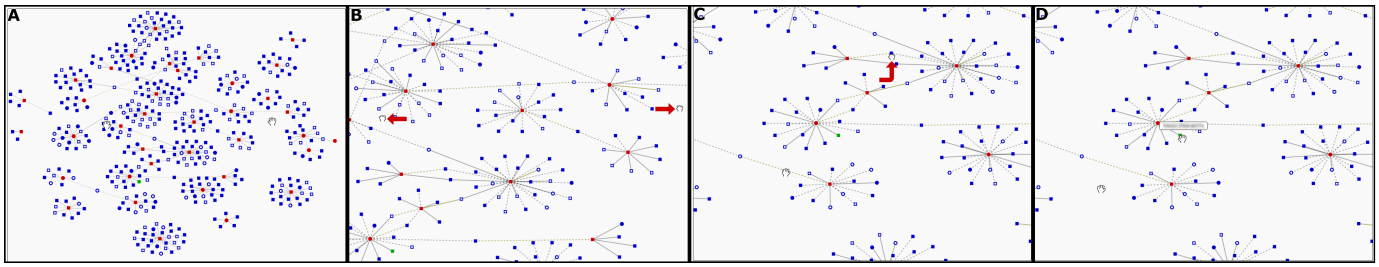

Fig. 1. Screenshots of a user performing a task in the system: (A) the initial screen of the application; (B) zooming into the graph to find a node; (C) translating the graph to put the green node in the center; (D) searching information of the green node.

***Abstract*—This paper presents a technique to provide natural interaction with public displays. The Microsoft Kinect is used to capture information about the user and an interpretation is made to identify what kind of interaction the user is trying to do. Gestures supported provide navigation, panning, and zooming, without any coupling devices on the user.**

***Keywords*-Gestural Interaction; Natural Interaction; HCI.**

I. Introduction

The use of large screens in public spaces such as airports, malls, bars and even squares in the streets informing upcoming events, promotions and other information that may be useful to users around it is becoming each day more common. Usually, these displays are not interactive and the user can only be updated with information. However, by providing interaction capabilities to the users, public displays have great potential for use, as has been seen in some recent works.

Traditional interaction devices are generally not suitable to interact with this kind of displays, especially when they are in public places, where there is no control over the environment. In these situations, it is suitable to provide interaction without any device additional.

Aiming to solve this problem, this paper presents a technique that uses Microsoft Kinect to recognize gestures of a user, providing a natural method of interaction without any direct use of devices. As a case study, an application that allows the interactive visualization of a social network was implemented and tested. This application runs in a browser, in which the Kinect was integrated to allow for user interaction.

II. Related Work

Researchers make use of various approaches to interact with large displays and the most common involves the use of touch-sensitive screens and devices that simulate such functionality [1]. Other strategies make use of mobile devices to interact with large screens [2], provide the development of new devices [3], or expand the functionality of existing devices, such as data gloves [4] for instance.

Even though all the strategies employed are able to solve the problem of interaction with large displays, there is scant research that does it without the use of any device by the user, as made by [5]. It is suitable especially when one wants a system capable of being conducted by different users in public areas.

This work wants to provide an intuitive interaction method that is low cost, robust, and capable of being used in actual uncontrolled environments.

III. Gestural Interaction

By developing an interaction method based on gestures in public areas, some problems must be taken into account, such as lighting, user identification, correct interpretation of user gestures, etc. Especially when it comes in a gestural interaction deviceless, to identify what the user is trying to do is not an easy task. In the approach used employing the Kinect to identify the user, the focus on the user is handled by the device, but other problems need to be addressed. In this work we try to identify some of these problems through the implementation of a case study.

A. Case Study

In the approach presented in this work, the user interacts with an information system using his hands and arms to browse and select virtual objects in a public display. The case study used is an application that runs in a Web browser, and shows a graph representing the academic social network of the Computer Science Graduate Program (PPGC) at UFRGS (see Figure 1). In this graph, nodes represent persons (professors and students), and edges indicate supervising relationships. In other words, the graph represents the genealogy of the PPGC.

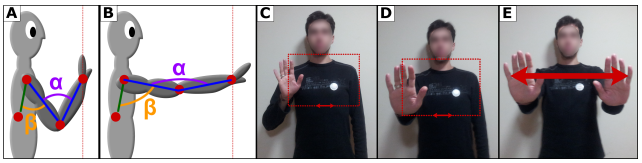

Fig. 2. Gestures in the system: (A) user with the arm close to their body indicates navigation; (B) user with outstretched arms indicates selection; (C) user performing navigation in the system; (D) user performing manipulation in the system; (E) user performing zoom-in and zoom-out, in a movement similar to those used in touch-screens.

To interpret the users' gestures, the system uses joints of their skeleton, easily identified by Kinect. For navigation, the system read the joints of the hands of the user and a selection is made when the user stretch his arm.

The depth information provided by Kinect is not accurated enough to identify this movement and, because of this, the response of the system became unstable. To solve this problem, the angles formed by the joints of the user's arm were used: between the arm and forearm (α), and between the arm and the side of the user (β). Thus, if α is greater than 110° and β is greater than 70°, the system interprets that the user is making a selection (Figure 2-B). Otherwise, the user is navigating (Figure 2-A).

In order to make these calculations, wrist, elbow and shoulder joints were used both on left and right sides of the skeleton, while the spine joints were used once. To provide robustness to the system, when the user reaches out completely toward the Kinect – in which case the device cannot correctly detect the skeletal joints – the depth information is used. This arrangement brings stability: when the user's hand is at the maximum distance that can be reached regarding the rest of his body (the dashed line in Figure 2-A,B).

The features developed in the case study are: nodes and graph manipulation, and a tooltip that shows the name of the person represented by the node being pointed by the cursor. The gestures used are: free movement of the hands along the body (Figure 2-C), for navigation; free movement with hands away from the body (Figure 2-D) for selection and manipulation; and approaching and moving apart both hands with the arms outstretched, to zoom in and out (Figure 2-E).

To integrate the Kinect with the browser, the client-server model was applied. The server is responsible for reading and interpreting the data obtained by Kinect and sends the information to the test application client. Two 3D coordinates are transmitted, corresponding to the positions of both hands, and two integers indicating that both hands are in navigation (0) or selection (1) state. With these data, the Web application can interpret which action is being taken: just moving the cursor around the graph to view information of the nodes; moving the entire graph or each individual node in the X and Y axes; or zooming the graph using both hands.

B. Implementation Details

The Microsoft official SDK was used to interpret data acquired with Kinect. The test application was developed in HTML5, JavaScript and PHP and runs on Web browsers. The library arbor.js (arborjs.org) was used to build the graph.

IV. Preliminary Results

Although the proposed solution does not work with extreme accuracy, using the zoom it is possible to manipulate individual nodes quite easily, even that some problems occur when the user stretches the whole arm and the system does not identify the desired selection. The solution presented for the selection mode, using angles of joints aided by depth information worked well enough, but still needs to be improved.

The solution presented behaved better than solutions that use only the depth information for selection, or those that use the absence of movement in which the user leaves their hand for a few seconds. However, it remains to be implemented a version that interprets the opening and closing movement of the hands to select objects. Details for this implementation are now being analyzed and the question is whether such identification does not adversely affect system performance.

V. Final Considerations

In this paper we present the first steps towards an implementation of a gestures-based interface for public displays. The Microsoft Kinect was used to identify the user and from this information, the selection and manipulation tasks were identified. The approach used runs in real-time and opens room for future improvements. A deep user study will be done to evaluate the robustness and usability of the system.

Acknowledgment

This work was partially supported by CNPq-Brazil under the project 311547/2011-7 and by Microsoft Brazil Interop Lab at UFRGS.

References

[1] G. Cohn, D. Morris, S. N. Patel, and D. S. Tan, "Your noise is my command: sensing gestures using the body as an antenna," in *Proceedings of the 2011 annual conference on Human factors in computing systems*, ser. CHI '11. New York, NY, USA: ACM, 2011, pp. 791–800. [Online]. Available: http://doi.acm.org/10.1145/1978942.1979058

[2] S. Boring, D. Baur, A. Butz, S. Gustafson, and P. Baudisch, "Touch projector: mobile interaction through video," in *Proceedings of the 28th international conference on Human factors in computing systems*, ser. CHI '10. New York, NY, USA: ACM, 2010, pp. 2287–2296. [Online]. Available: http://doi.acm.org/10.1145/1753326.1753671

[3] L. Aguerreche, T. Duval, and A. Lécuyer, "Reconfigurable tangible devices for 3d virtual object manipulation by single or multiple users," in *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology*, ser. VRST '10. New York, NY, USA: ACM, 2010, pp. 227–230. [Online]. Available: http://doi.acm.org/10.1145/1889863.1889913

[4] M. Moehring and B. Froehlich, "Effective manipulation of virtual objects within arm's reach," in *Virtual Reality Conference (VR), 2011 IEEE*, march 2011, pp. 131 –138.

[5] M. Nancel, J. Wagner, E. Pietriga, O. Chapuis, and W. Mackay, "Mid-air pan-and-zoom on wall-sized displays," in *Proceedings of the 2011 annual conference on Human factors in computing systems*, ser. CHI '11. New York, NY, USA: ACM, 2011, pp. 177–186. [Online]. Available: http://doi.acm.org/10.1145/1978942.1978969