Multi-user Pointing and Gesture Interaction for Large Screen Using Infrared Emitters and Accelerometers

Leonardo Angelini, Maurizio Caon, Stefano Carrino,
Omar Abou Khaled, and Elena Mugellini

College of Engineering and Architecture of Fribourg, Switzerland
{Leonardo.Angelini,Maurizio.Caon,Stefano.Carrino,
Omar.AbouKhaled,Elena.Mugellini}@Hefr.ch

Abstract. This paper presents PlusControl, a novel multi-user interaction system for cooperative work with large screen. This system is designed for a use with economic deictic and control gestures in air and it allows free mobility in the environment to the users. PlusControl consists in light worn devices with infrared emitters and Bluetooth accelerometers. In this paper the architecture of the system is presented. A prototype has been developed in order to test and evaluate the system performances. Results show that PlusControl is a valuable tool in cooperative scenarios.

Keywords: human computer interaction, large screen, gesture recognition, visual tracking, computer supported cooperative work, economic gestures.

1 Introduction

In 1980, A. Bolt introduced his Put-That-There system to the scientific community [1]. He started an important revolution in the Human-Computer Interaction area; in fact, he stated that the machine should understand the human language and not the contrary, making interaction between human and machine more natural. The natural way of interaction between human beings is based on speech and gestures. Therefore, in order to make human computer interaction more natural, a system should get closer to these forms of multimodal communication [2]. Unfortunately the voice commands are not convenient in any situation. For example, in a conference scenario there could be some troubles with the speech of the people making a presentation. In fact some words could be incorrectly understood by the computer as a voice command. Adding a gesture as trigger for the speech command could be a solution but such a way of interaction introduces a break in the presentation. Our aim, instead, is to make the interaction and the presentation two seamless and synergic experiences. Therefore a system that involves exclusively gesture recognition should be preferable.

The support for multi-user interaction can be really important in areas that involve cooperative work such as for instance educational application. An example of a cooperative work application can be found in [3], in which new devices (e.g. IntelliPen, a laser tracked pen for direct manipulation of objects on the projection wall) and interaction modalities (e.g. gestures) have simplified and accelerated the engineering and design development phases. Moreover, single-display groupware

J.A. Jacko (Ed.): Human-Computer Interaction, Part II, HCII 2011, LNCS 6762, pp. 185–193, 2011.

systems have been discovered as an important tool to make students more motivated, more involved in the class and more socially interactive [4]. At present, those systems utilize several mice for the interaction but a free air system should be more comfortable. Ying Tin and Randall Davis stated that interfaces closely conformable to the way people naturally interact would have the potential to lower the users' cognitive load, allowing them to concentrate on the decision-making task. This is a feature that should be precious in an educational context where the users are supposed to interact focused on contents. The same concept has been applied in the creation of the USAR (urban search and rescue) system to help people in organizing and coordinating rescue tasks in a high pressure context, like aiding victims in the aftermath of a disaster [5].

The system presented in this paper is based on the recognition of pointing and control gestures performed in air granting simultaneous interaction of multiple users on a large screen. PlusControl is designed for natural and economic interaction, facilitating its learnability.

Reaching large screen with small movements is not easy without losing precision, but the developed system exploits the speed of hand movements to grant higher precision and smaller displacement with slow gestures, and larger displacements and lower precision of the cursor with quick gestures. Moreover the interaction has been designed to perform gestures which require small movements in order to make a less tiring experience [6].

This system combines the advantages of an easy-to-use pointing system with a natural gesture language for commands.

In order to provide a clear research context for PlusControl, the related works are analysed in Section 2. In Section 3, an overview of the PlusControl architecture is presented. In Sections 4 and 5, the testing methodology and the obtained results are shown. Section 6 concludes the paper and presents future works.

2 Related Works

Interaction with large screens is a demanding area for research because of the wide diffusion of very large displays and projectors. In fact realizing an interface for these applications leads to many serious technical problems. For example, the large size of these screens can cause users to lose track of the mouse pointer and it can make really difficult reaching distant contents [7]. To cope with these problems, many systems, based on different technologies, have been developed. Some systems have touch interfaces (e.g. [8]) but they require the users to remain near to the displays. Therefore, to grant free mobility to the users in the environment while maintaining the possibility to continue the interaction with the system, many researchers adopted different solutions for pointing. Most of them have chosen laser pointers as cursors [9], [10], [11]. This technique is based on computer vision: the laser dot on the screen is captured with a camera and the image is processed to obtain the cursor information. Using only the pointing system, it is possible to control an interface but it is necessary to introduce complex solutions to realize even the simpler commands like clicking. For example, in [12], a LED source is tracked by a camera for pointing; the clicking action is obtained by the double crossing interaction. In the article [13] the authors put a button on the laser pointer to click. Jiang et al. made a pointer handheld device combining a camera with the Logitech Cordless Presenter [14]. Clicking is an essential feature for most applications, but a pointing system could be augmented by gestures to execute more actions with free mobility, granting a seamless experience. In the gesture recognition domain, related works are innumerable. There are many technologies that allow movements capturing; among them, accelerometers are really quoted by researchers since they are already present in several devices, have a low price and achieve excellent performances [15], [16], [17], [18], [19].

Many systems allow pointing and clicking granting free mobility; others implement gesture recognition for a natural interaction. PlusControl is a novel system that combines these two features adding the possibility of multi-user interaction for cooperative work.

3 PlusControl Architecture

The PlusControl system is designed to support concurrent multi-user interaction on a large screen. In order to allow to the users a seamless and natural cooperation, the system showed in Fig. 1 has been developed. The interaction system is composed of two accelerometers and an infrared emitter for each person, which can be worn on a finger or held in the hand. The infrared emitter is tracked with two IR cams. The two cameras are horizontally parallel; they are positioned turned towards the users and under the large screen. The small control movements allow using only a part of the range of vision of the cameras. Moreover, two parallel cameras grant a wider area of interaction, thus a multi-user control is possible. Accelerometers are used to combine the cursor movement with both simple gestures (e.g. press click, release click, simple click) and more complex gestures (e.g. user recognition, previous slide, next slide).

The algorithm that calculates the displacement of the cursor using data from the IR emitters tracking has been improved with a distance factor correction based on stereo vision triangulation. In fact, using two IR cameras it is possible to calculate the approximated distance of the user by measuring the horizontal disparity of the two blobs tracked from each of the two cameras. The correction factor makes the interaction not dependent from the distance of the user from the cameras. This approach grants small movements of the hand also at large distances, improving the usability.

The main difficulty in a multi-user environment is to associate the correct tracked blob to each user. The user recognition is done with a particular gesture that allows matching the accelerometer data with the tracked blob data. A quick up-down movement has been chosen as gesture for the user recognition (assigning the accelerometers data to the tracked blob) when his emitter enters the range of vision of the camera.

Using the distance correction system, the blobs to be assigned to each user are two. As a consequence even if one of the two cameras loses its blob, the system is still able to move the cursor using the other blob but the distance of the user will not be calculated. Until one of the two blobs is tracked, the system is able to re-associate the second blob when it reaches again the field of vision of the cameras. If both the blobs are lost, the user should do again the gesture to be recognized from the system (i.e. the quick up-down movement).

In addition working with infrared light grants better robustness against light changes.

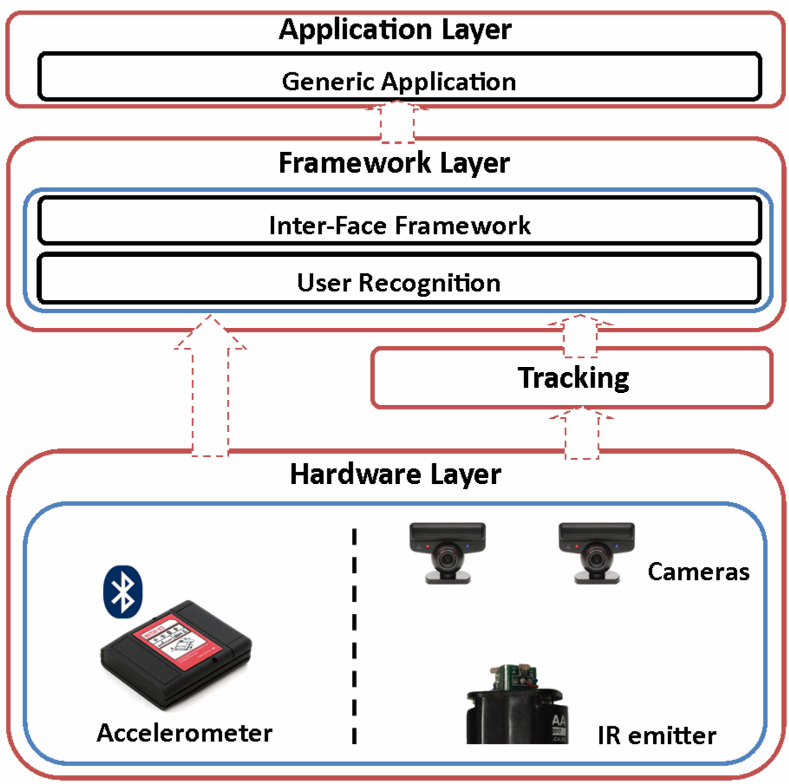

Fig. 1. PlusControl system architecture

4 Prototype

As shown in Fig. 1, the system has been integrated in an innovative framework, called Inter-Face that manages the creation of interactive surfaces supporting multi-user interaction [20]. Standard applications in a Windows environment, in fact, do not allow the use of multiple cursors. The Inter-Face Framework, instead, allows multiple inputs (i.e. several users interacting at the same time on the same surface) and combines multiple technologies (i.e. it integrates different sensing technologies such as acoustic, optic and RFID in order to support multimodal interaction). The PlusControl system has been integrated in the Inter-Face Framework as a new interaction modality.

The PlusControl system can recognize different gestures. Turning the left wrist counter-clockwise performs the press-click action. Turning the left wrist clockwise releases the click. While the press-click state is activated, turning the right wrist allows a rotation of 90° of the selected object in the application. This system allows configuring the devices symmetrically opposite to the previous explication; therefore it provides the possibility to perform gestures specularly. A quick horizontal movement in opposite directions of the hands allows zooming in or out the object.

The quick up-down gesture is used to recognize the user when the pointer is lost, exploiting the correlation of information from the tracked blob and the accelerometer of the arm in which the infrared emitter led is worn.

Two prototypes of the infrared emitter have been developed. The first one can be worn on a finger (Fig. 2 a)), while the other one can be held in the hand (Fig. 2 b)). Both prototypes have three IR LEDs placed on a triangular support; they allow tracking the IR source even with high inclination angles between the axis of the camera and the axis of the emitter. Both emitters are powered by two 1.5V AA batteries, which allow an operative time of about 48 hours. The emitters are tracked by two filtered Sony PS3 Eye cameras with a resolution of 640x480 pixels at 60 frames per second. Two Bluetooth SparkFun WiTilt V3.0 accelerometers are worn by the user on the wrists and they operate at a frequency of 50 Hz.

Fig. 2. Prototypes of the body worn infrared emitter (*left*) and of the handheld infrared emitter (*right*)

5 Testing Methodology

Some preliminary tests have been performed in order to assess the effectiveness and the usability of the system, also in a cooperative work scenario, by subjective and objective evaluations. We have invited 10 students (8 boys and 2 girls) to evaluate the system. We have created three applications to test this system and its many features.

The test was composed of 4 steps.

The step 1 consisted in solving a puzzle (a 1024x768 pixels image) utilizing in a first time the mouse and after our system. The puzzle was composed of 16 pieces (each one of 256x192 pixels) that were randomly distributed on a 4x4 table, as shown in Fig. 3 b). The users had to put the pieces in the right table cells to visualize the original image. The screen was projected on a wall with a 1024x768 pixels resolution. We gave to the users the time to understand the image to compose and to try the control system. After this phase, we asked them to solve the puzzle and we recorded the time.

The step 2 was similar to the first one. The image to be recomposed had a dimension of 1280x1024 pixels and it was split in 16 pieces (each one of 320x256 pixels). The application was displayed with two monitors with resolution at 1680x1050 pixels. The interactive surface of the application was of 3280x1024 pixels. The 4x4 table was in the left side of the screen and the images randomly distributed in the right side, as shown in Fig. 3 a). The users had to put the pieces in the right table cells to visualize the original image. The users have utilized before the mouse and after the PlusControl system while we were recording the time.

The step 3 consists of solving the puzzle of the step 1 and of the step 2 in cooperation with another user. We recorded the time.

In the step 4 the users had to manipulate the images in an application using the gestures presented in the Architecture paragraph. The visual interface of this application is shown in Fig. 3 c).

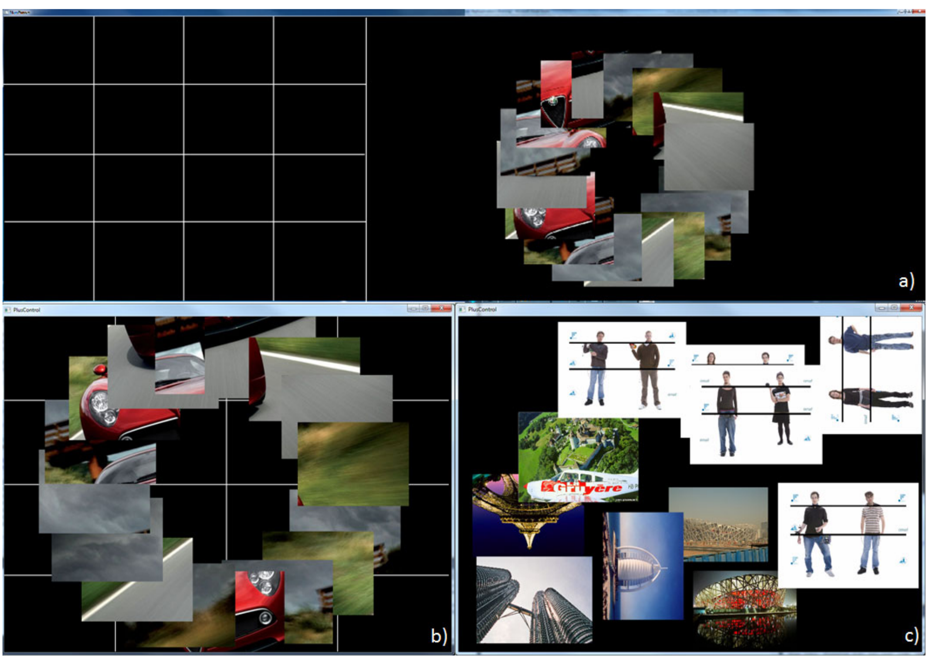

Fig. 3. a) The puzzle application for the test with the 3280x1024 px screen. b) The puzzle application for the test with the 1024x768 px screen. c) The application for the free manipulation of the images.

Finally, the participants had to fill in a questionnaire, rating the system features according to a 7-point Likert scale. The evaluation parameters were comfort, precision, learnability and cooperation. Comfort considers how tired the user was after the test session and how easy was reaching all the parts of the screen. Precision is strictly related to the easiness and accuracy of the cursor and commands control. Learnability refers to how much the user, which utilized PlusControl for the first time, got acquainted in performing pointing action and gestures. Cooperation measures how much a participant has appreciated collaborate with another person and if he had difficulties to accomplish the given task. The questionnaire included a space for suggestions and comments.

6 Results

Tests have been executed following the methodology presented in Section 5. Results of these tests are discussed in this Section. Fig. 4 presents average solving time and relative standard deviation of steps 1, 2 and 3. In the graphic on the left are presented the results obtained with the 1024x768 pixels screen projected on the wall. The graphic on the right shows the results in the large displays scenario. Using the PlusControl system, the participants achieved a solving time that is 24% and 33% higher than utilizing the mouse, respectively to the step 1 and 2. Solving times have been drastically reduced in cooperation scenarios. Results of the subjective evaluation are presented in **Table 1.**. Users expressed their appreciation about easiness of use and pointer reactivity. On the other hand they have suggested improving the reactivity of the gesture recognition subsystem.

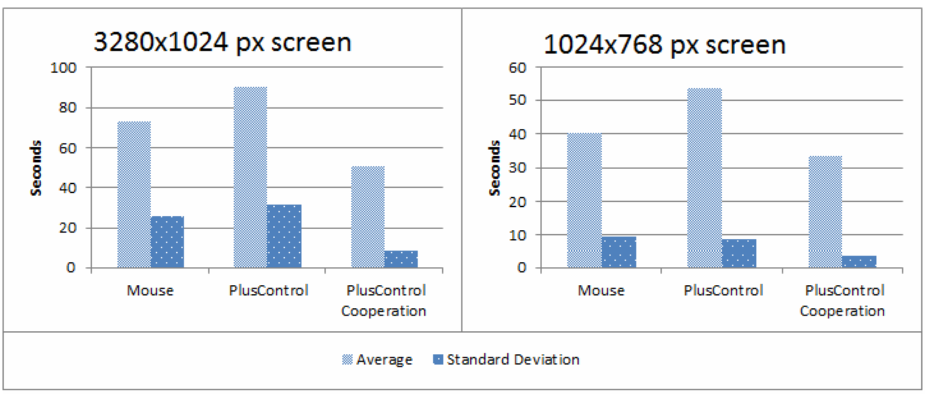

Fig. 4. Average time (and standard deviation) for puzzle completion on 3280x1024 px screen (*left*) and 1024x768 px screen (*right*)

Table 1. Results of the subjective evaluation of PlusControl main features according to a 7-point Likert scale

Features	Average	Standard Deviation
Comfort	5.7	0.7
Precision	5.5	0.8
Learnability	5.3	0.7
Cooperation	6.2	0.6

7 Conclusions and Future Works

In this article a novel interaction system that allows to several users to cooperate on a shared large screen has been presented. PlusControl combines an infrared pointing system with the recognition of several natural gestures using accelerometers. The test showed that users appreciated the system and that it can be very useful in cooperative scenarios. In the future works the system will be tested with more than two users at the same time, in order to infer the maximum number of users working contemporaneously according to technology constraints. Moreover we will improve the performances of the system by developing a specific device which will integrate the triaxial accelerometer and the infrared emitter used so far, plus a gyroscope.

References

1. Bolt, R.A.: "Put-that-there": Voice and Gesture at the Graphics Interface. ACM SIGGRAPH Computer Graphics 14, 262–270 (1980)
2. Krahnstoever, N., Kettebekov, S., Yeasin, M., Sharma, R.: A real-time framework for natural multimodal interaction with large screen displays. IEEE Comput. Soc., Los Alamitos (2002)
3. Kela, J., Korpipää, P., Mäntyjärvi, J., Kallio, S., Savino, G., Jozzo, L., Marca, S.D.: Accelerometer-based gesture control for a design environment. Personal and Ubiquitous Computing 10, 285–299 (2005)
4. Amershi, S., Morris, M.R., Moraveji, N., Balakrishnan, R., Toyama, K.: Multiple mouse text entry for single-display groupware. ACM Press, New York (2010)
5. Yin, Y., Davis, R.: Toward Natural Interaction in the Real World: Real-time Gesture Recognition. ACM Press, New York (2010)
6. Shintani, K., Mashita, T., Kiyokawa, K., Takemura, H.: Evaluation of a Pointing Interface for a Large Screen Based on Regression Model with Image Features. IEEE, Los Alamitos (2010)
7. Baudisch, P.: Interacting with Large Displays. Computer 39, 96–97 (2006)
8. Morrison, G.D.: A camera-based input device for large interactive displays. IEEE Computer Graphics and Applications 25, 52–57 (2005)
9. Fukuchi, K.: A laser pointer/Laser trails tracking system for visual performance. In: Costabile, M.F., Paternó, F. (eds.) INTERACT 2005. LNCS, vol. 3585, pp. 1050–1053. Springer, Heidelberg (2005)
10. Zhang, L., Shi, Y., Chen, B.: NALP: Navigating Assistant for Large Display Presentation Using Laser Pointer. In: First International Conference on Advances in Computer-Human Interaction, pp. 39–44 (2008)
11. Davis, J., Chen, X.: Lumipoint: multi-user laser-based interaction on large tiled displays. Displays 23, 205–211 (2002)
12. Nakamura, T., Takahashi, S., Tanaka, J.: Double-crossing: A new interaction technique for hand gesture interfaces. In: Lee, S., Choo, H., Ha, S., Shin, I.C. (eds.) APCHI 2008. LNCS, vol. 5068, pp. 292–300. Springer, Heidelberg (2008)
13. Oh, J.Y., Stuerzlinger, W.: Laser pointers as collaborative pointing devices. Graphics Interface, pp. 141–150, Citeseer (2002)
14. Jiang, H., Ofek, E., Moraveji, N., Shi, Y.: Direct pointer. ACM Press, New York (2006)

15. Barbieri, R., Farella, E., Benini, L., Ricco, B., Acquaviva, A.: A low-power motion capture system with integrated accelerometers gesture recognition applications. In: 2003 International Symposium on VLSI Technology, Systems and Applications. Proceedings of Technical Papers (IEEE Cat. No.03TH8672), pp. 418–423 (2004)
16. Liu, J., Wang, Z., Zhong, L., Wickramasuriya, J., Vasudevan, V.: uWave: Accelerometer-based personalized gesture recognition and its applications. In: IEEE International Conference on Pervasive Computing and Communications, pp. 1–9 (2009)
17. Mäntyjärvi, J., Kela, J., Korpipää, P., Kallio, S.: Enabling fast and effortless customisation in accelerometer based gesture interaction. In: Proceedings of the 3rd International Conference on Mobile and Ubiquitous Multimedia - MUM 2004, pp. 25–31 (2004)
18. Fu, L.C., Fellow, I.: Gesture stroke recognition using computer vision and linear accelerometer. In: 8th IEEE International Conference on Automatic Face & Gesture Recognition, pp. 1–6 (2008)
19. Schlömer, T., Poppinga, B., Henze, N., Boll, S.: Gesture recognition with a Wii controller. In: Proceedings of the 2nd International Conference on Tangible and Embedded Interaction - TEI 2008 (November 2008)
20. Mugellini, E., Abou Khaled, O., Pierroz, S., Carrino, S., Chabbi Drissi, H.: Generic framework for transforming everyday objects into interactive surfaces. In: Jacko, J.A. (ed.) HCI International 2009. LNCS, vol. 5612, pp. 473–482. Springer, Heidelberg (2009)