"di Piazza in Piazza": Reimagining Cultural Specific Interactions for People-centered Exhibitions

Silvia Mirri, Paola Salomoni,
Alberto Pizzinelli, Marco Roccetti
Department of Computer Science and Engineering
Università di Bologna
Bologna, Italy
{silvia.mirri, paola.salomoni, alberto.pizzinelli,
marco.roccetti}@unibo.it

Claudio E. Palazzi
Department of Mathematics
Università di Padova
Padua, Italy
cpalazzi@math.unipd.it

***Abstract*— One of the most interesting innovations introduced by modern game console technologies has been that of new interaction systems which allow a large amount of people to interact all together with cultural artifacts and exhibits. Thank to them, human beings can interact with, manipulate and in some circumstances even hack digital contents, playing an active role and becoming part of cultural performance exhibitions. In this paper we present and discuss our recent experience in designing, implementing and putting in practice a hands-free interactive technology which lets visitors explore a hyper-multimedia system, playing a role of those who are able to transform historical audio clips and videos into a piece of public culture which needs to be lived together on a public place. The hands-free system in use is a part of a wider project (called "di Piazza in Piazza"), which comes with the initiatives of the University of Bologna, currently chosen to be displayed at the 2015 Universal Exposition in Milan (EXPO 2015).**

Keywords—Multimedia and Interactive Systems, Exhibitions, Gesture Recognition, Hands-Free Technology

I. INTRODUCTION

From the seventh generation of game consoles (with the Nintendo Wii launch) to the announced Oculus Touch (available in the first quarter of 2016), the trend in terms of interface mechanisms is to let the user interact hands-free or by using very small and ergonomic devices [1]. This means that users can interact with technology-augmented objects, while performing intuitive and natural gesture with their hands and their entire body [2]. Such an interaction mechanism could be successfully exploited in several and different fields, with the aim of equipping users with richer and more complex interfaces to digital contents. One of these fields is represented by the technology-based exhibitions and performances, which can be located in cultural public places, such as museums, libraries, fairs, expositions, etc. Some projects [2, 3] are devoted to go further in the direction of equipping visitors of exhibitions with the opportunity to enjoy and explore (and sometimes manipulate and hack) multimedia contents with a more active role than simply passively listen and watch them. The use of modern game consoles can provide great benefit in this sense, reaching such a goal with simple and low-cost technologies and equipment.

In such a context, we want to describe our experience in designing, implementing and putting in practice a hyper-multimedia system within the initiatives carried on by the University of Bologna together with the Municipality of Bologna on the occasion of the Universal Exposition (EXPO2015), which is hosted in Milan in 2015. The EXPO2015 theme is *Feeding the planet, energy for life*. The city of Bologna is renowned for its culinary tradition and the food, food markets and marketplaces are the topic of several researches at the University, in different disciplines, including humanities and social ones (from history to arts, from food to life sciences). Hence, the idea was to collect the main research results in such fields and to expose them to the citizens. The project has been named "di Piazza in Piazza – viaggio nella cultura alimentare" (which literally means "From Square to Square – a journey through the food culture"), because in Italy squares are the main marketplaces since the Roman Empire. The project aimed to host such an exhibition not only in Bologna, but also in the other University of Bologna campuses (in the cities of Cesena, Rimini, Ravenna, Forlì and Imola). Instead of simply basing our system on a passive video and audio narration, we decided to design and implement a mechanism which lets the user interact with the multi and hyper media system by means of his/her hands, exploring and manipulating the contents with a more active role. At the time of this writing, the di Piazza in Piazza exhibitions in Salaborsa Library (Bologna) and in the Museum of the City (Rimini) are yet concluded. We are currently setting up the exhibitions which will be hosted in Cesena (Malatestiana Library) and in Ravenna (Palazzo Rasponi Dalle Teste).

In order to implement such a system, we had to face with different scientific and technological challenges. One of them was devising the most appropriate techniques and algorithms able to automatically recognize a set of hand gestures, with simple and low-cost hardware. Moreover, putting in practice such an interactive mechanism had to deal with the specificities of the contexts and locations of use (museums and libraries, and other public cultural places, where several visitors could interact with and enjoy the system), taking into account limiting technical constrains, environmental contexts (e.g. noise and light conditions) and different set up configurations.

The remainder of this paper is organized as follows. In Section II, we present some projects similar to our work, which have inspired it. A description of main design issues and involved technologies is given in Section III. Section IV presents the most challenging aspects of the implementation, together with a short overview of the contents organization. Section V reports about the first "di Piazza in Piazza"

exhibitions, which were successfully concluded. We finally conclude the paper with Section VI.

II. Background and Related Work

Several works have been done in the research field of advanced gesture-based human-computer interaction interfaces. In this paper, we focus on the on the research results that are more closed to ours, in particular in terms of goals and technological issues. Here, we refer to three of those works which exploited advanced digital recognition technologies in the contexts of artistic and exhibitive performances, and which inspired our work.

An interesting experience has been done at IRCAM, as described in [3]. The project is related to the process of learning music theory and it involves both hardware and software components (accelerometers and gyroscopes and a gesture following and recognition system). Its main goal is to equip a user with a tool useful in the process of learning how to conduct a music orchestra. The system checks users' gestures and provides feedbacks about their correctness. In this system the interaction is based on the use of specific hardware and sensors, while the idea at the basis of our project was to let the user interact with our application hands-free, limiting as much as possible the involvement of additional technologies.

Some other works and exhibitions have been based on the use of the Nintendo Wiimote in a non-traditional way. In particular, the original architecture of the Nintendo console uses the remote controller to receive red light signals from a sensor bar source and to transmit some kinetics parameters to the system via Bluetooth, while the proposed approach inverts the flow of information. In fact, instead of using the remote controller as a receiver, these works exploited it as a sensor system, recognizing red light emitters from a device. In particular, [4] proposed a virtual aerosol bomb, which can interact with the most common graphics applications (e.g. The Gimp, Photoshop, MSPaint). On the contrary of our project, these works imposed the use of additional input devices.

Another interesting experience was related to the Italian food and city of Bologna and was exposed to the Universal Exposition (Shanghai 2010): Tortellino X-Perience [1, 2] is a multimedia system designed to let the users learn to prepare tortellini, a home-made pasta, typical from the city of Bologna. In particular, Tortellino X-Perience is a sort of interactive storytelling, where the player exploits an explanation and then tries to repeat the correct gestures needed to model tortellini, by mimicking a real cook. The system tracks the users' gestures and provides feedback about their correctness. The users can interact with the system with just their hands, without any additional device or sensor, and only a camera has been used to capture and track users' movements. This project is the closest to our one and has inspired and driven our work.

III. Design Issues

This project was born as an initiative of the University of Bologna together with the Municipality of Bologna on the occasion of the Universal Exposition, which is hosted by Milan in 2015. The theme of the EXPO is *Feeding the planet, energy for life*. The city of Bologna is renowned for its culinary tradition and the food, food markets and marketplaces are the topics of a wealth of research at the University, in different disciplines, including humanities and social ones. Hence, the idea was to collect the main results in such fields and to expose them to the citizens. The project goes under the name of "di Piazza in Piazza – viaggio nella cultura alimentare", which literally means "from Square to Square – a journey through the food culture". In fact, in Italy, squares are the main marketplaces since the Roman Empire. The project aimed to expose such an exhibition in Bologna and in the other University of Bologna campuses (in the cities of Cesena, Rimini, Ravenna, Forlì and Imola).

The project involved more than 30 content authors (professors and researchers), who provided the results of their studies as pictures, videos, audios, and textual documents, in form of raw materials. In particular, ten main themes have been identified, so as to group all the gathered contents by: *Women and food*, *Bread*, *Pork*, *Fashion and food*, *Wine*, *Fruits and vegetables*, *Markets*, *Fish and salt*, *Harbors and means of transport*, *Cook, kitchen and recipes*. Summing up, more than 2,000 files related to such raw materials have been collected, for a total amount of more than 30 GB. The large amount of content together with the variety of location characteristics have driven the design phase, pointing out the following needs: (i) contents have to be different from a passive video and audio narration; hence the technologies at the basis of the installation have to let the user explore the content with a more active role; (ii) technologies and contents (in terms of shapes and formats) have to be: portable, replicable, with limited constrains, low-cost, easy to be used and installed.

In order to respond to these needs, we chosen to exploit the Kinect, an advanced video camera and software that supports human body recognition. The Kinect sensor [5] is a flat black box that sits on a small base, which has been designed to be positioned lengthwise above or below the video display. Main hardware components are the following ones: a video camera, a depth sensor and a multi-array microphone. The depth sensor is composed by an infrared laser projector and by a monochrome CMOS (Complementary metal–oxide–semiconductor) sensor and it can capture video data in 3D, regardless of the lighting conditions. The multi-array microphone is composed by four microphones that can isolate users' voices from the noise in the room. The Kinect sensor works in two main steps [6]: (i) a depth map is created by exploiting the light analysis the infrared laser projector constructs; (ii) the positions of players are computed by exploiting tracking algorithms. All the body parts of the players in the field-of-view of the Kinect camera are recognized and 3D models are constructed. Joints are applied to such a 3D model (corresponding to specific body parts), reconstructing a sort of skeleton of the player [6]. We exploited such a sensors system directly integrated with a laptop, without the Microsoft Xbox Console. This allows the users interact with the hyper-multimedia contents with their body. In particular, we decided to let the user explore the content by means of their hands, while performing intuitive and natural gestures and emphasizing food handling.

Since we had no detail about the locations of the exhibition and about the kind of display which will be used in the

different hosts, we planned different set up configurations, with the aim of limiting technical constrains (e.g. connection requirements) and enforcing the characteristics of portability, replicability, easiness of use and of being installed and configured. In particular, we identified two main groups of configurations, on the basis of the display, which can be a led-wall (Figure 1) or a projection on a wall or on a screen (Figure 2). The configurations we defined are based on the Kinect position (which can be placed above or below the video display) and take into account physical constrains, such as the distance between the Kinect and the user (approximately between 1.8 and 2.4 meters) and the distance between the video display and the user (at least 3.0 meters if the display in use is the led-wall). Such configurations have driven the design, the implementation and the test phases and have played an important role in defining the locations of the exhibition. Figure 1 shows a set up configuration we planned on the use of a led-wall, which could be used in wide spaces with a strong lighting (such as the Salaborsa in Bologna), while Figure 2 depicts a set up configuration based on the use of a projector, which could be used in smaller places with any other lower lighting, ranging to penumbra (such as the Museum of the City of Rimini).

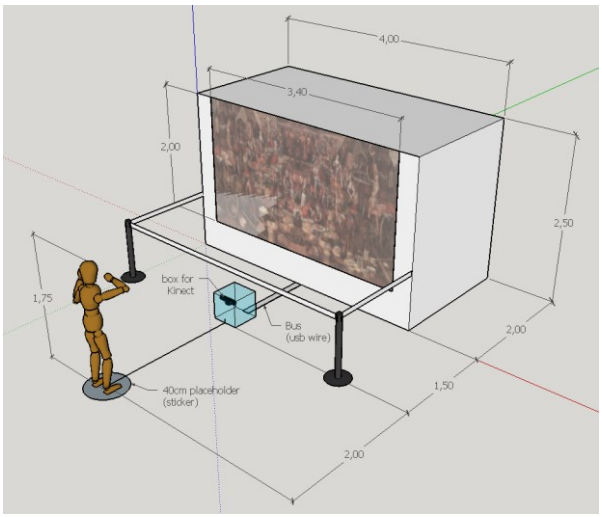

Figure 1 – Set up configuration with a led-wall.

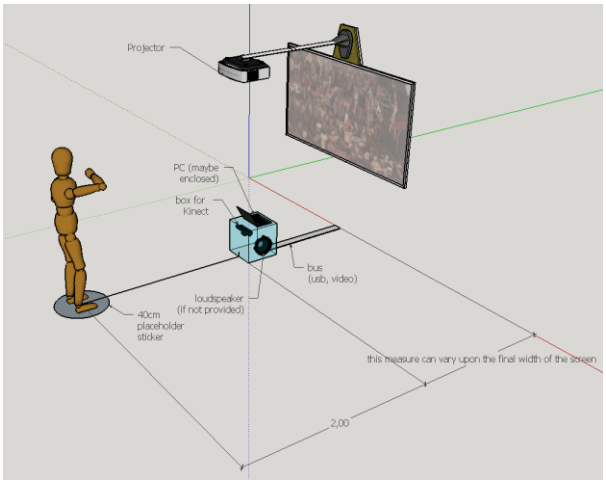

Figure 2 - Set up configuration with a projection.

IV. Implementation

This section presents the most relevant issues related to the implementation and the putting in practice of the di Piazza in Piazza system, which is based on two main aspects: the content authoring and editing (subsection A) and the gesture control set up (subsection B).

A. Content Authoring and Editing

As we have already mentioned, several colleagues of us offered the results of their studies about the food and the markets in different disciplines (from history to arts, from food to life sciences). Such results have been selected and post-produced in a multi and hyper-media which can be navigated by means of a browser. In particular, the di Piazza in Piazza contents have been structured as follows:

- An introductory video, which resumes the aims, the topics and the actors of the project.
- A tag-cloud, which lets the users browse the ten main topics (related to the theme listed in the previous Section). A screenshot is shown in Figure 3.
- Ten scenes, one for each main topic, which let the users browse among deepening contents.
- Deepening contents (from 5 to 6 for each main topic). Most of them are short videos about specific aspects of the scene, but they also include QR-codes which let the users explore additional contents by means of their smartphones.

The whole resulting multi and hyper-media includes and synchronizes videos, images and audios. It has been developed by exploiting HTML5, CSS3 and JavaScript, with the aim of responding to the needs described in the previous Section.

We have created a post-production mechanism which is based on XSL transformations. It can be configured and set up in a quick and simple way so as to adapt the multimedia system to the hardware configuration in use. This lets the authors add in an easy way new contents (in terms of scenes and of deepening ones) and it allows the use of different contents formats, limiting the need of technical support.

Each scene is configured by means of an XML file (referring to the media files related to the corresponding deepening contents) and the XSL transformation automatically produces the corresponding set of HTML pages [7].

Figure 3 - A screenshot of the tag-cloud

B. Gesture Control Set Up

Another interesting part of the project is the users' interaction with the di Piazza in Piazza content by means of the Kinect sensors system, which recognizes and tracks user's movements. In order to do that, we have implemented specific procedures, taking into account the following goals:

- identifying one user among the others in the Kinect camera field of view (the one who drives the navigation among the content);
- identifying one user's hand;
- linking the hand with the mouse pointer;
- identifying a specific gesture as the click event;
- letting the user change the controlling hand (from the right to the left one and vice versa).

In order to reach these goals, we have exploited Microsoft Kinect SDK and we have worked with its Natural User Interface (NUI) library (developed in C#). Such a library lets the developer access all the data coming from the Kinect sensors, including the tracking ones and the users' skeleton 3D models. In particular, we have handled the skeleton of the users, player's hands and the pointer, and the pointer and interface, as described in the following.

The exhibition has been designed to be located in museums, libraries or other public and open spaces. This means that more people could be identified at the same time in the field-of-view of the Kinect camera. We have managed this situation by capturing and activating the first skeleton and letting such user interact with the system. Thanks to the NUI, we check if there are one or more skeletons tracked by the Kinect. The data about all the skeletons and their joints are saved in an array. We implemented a procedure which handles users' priority in a sort of FIFO management: when the player leaves the application and his/her skeleton is no more in field-of-view of the Kinect camera, then the second skeleton acquires the Kinect control and the corresponding user can interact with the application, and so on.

In order to let the user interact with the application by means of his/her hands, we have identified the corresponding joints in the active skeleton and linked them with the mouse pointer. Both hands are identified and the foreground one is associated to the mouse pointer. This means that the users can indifferently use the right or the left one and they can also changes the hand which control the mouse pointer during the interaction with the di Piazza in Piazza application. Hence, the mouse pointer is directly associated to the hand: the Kinect sensor tracks the hand movement and adequately moves the mouse pointer. By exploiting the system libraries, we have associated a specific gesture to the click event. When the click gesture is recognized by the Kinect sensor, then a mouse left button click is emulated. Clicking on specific elements in the di Piazza in Piazza application let the user navigate the contents. A chunk of the code of the Hands and Pointer Handler is shown in Code 1. **`SendClick`** emulates the click, by exploiting **`SendDown`** and **`SendUp`**, while **`MoveMouse`** is the procedure in charge with the mouse movement, where the **`p`** variable is a **`Point type`**, which memorizes the controlling hand coordinates, relative to the display in use.

The user interface has been implemented by using the Windows Presentation Foundation (WPF), a .NET framework library which is devoted to the GUI development in Windows environments. Thanks to this library, the graphical layer of the application which interacts with the Kinect sensor has been added, over the display one. Such an additional layer is transparent and this lets the user exploit the application over the objects and the elements shows by the PC, directly interacting with them by means of the mouse pointer.

The mouse pointer is highlight by means of a red circle (see Figure 3). When the user chooses to click an element, he/she moves the controlling hand since the mouse pointer (and the red circle) reaches such an element, which is highlighted thanks to the **`onmouseover`** event; when the red circle arrives at a clickable element, then the circle size decreases and when it is just a red dot then the click is definitively activated.

V. Di Piazza in Piazza Ehxibitions

The di Piazza in Piazza exhibition has been hosted in Bologna (Salaborsa library) and in Rimini (Museum of the City). The exhibition in Salaborsa took place in May 2015 for three weeks; since a led-wall was used, the configuration set up was the one shown in Figure 1. The exhibition took place at the ground floor of Salaborsa, which is a wide indoor square with a very high ceiling (see Figure 4 and 5). The led-wall size was 3.5x2 meters, with a 1366x768 screen resolution. The Kinect sensor was placed in a box in front of the led-wall, 1.5 meters distant from the screen. The best position where the users can interact with the application was pointed out on the floor (2 meters from the Kinect sensor) by a sticker with the di Piazza in Piazza logo. Speakers were placed over that position, so as to let the users better enjoy the audio in such a wide hall (see the transparent bell over the users in Figures 4 and 5). The average affluence was about 200 persons per day during the week and 300 persons per day in the weekends, ranging from schoolchildren to elderly people, including tourists and Bologna citizens.

```
public class MouseController
{
  […]
  internal static void SendClick()
  {
      sendDown();
      sendup();
  }
  internal static void SendUp()
  {
      Mouse_event(MOUSEEVENTF_LEFTUP, 0, 0, 0,
                  new System.IntPtr());
  }
  internal static void SendDown()
  {
      Mouse_event(MOUSEEVENTF_LEFTDOWN, 0, 0, 0,
                  new System.IntPtr());
   }
  internal static void MoveMouse(Point p)
  {
      SetCursorPos((int)p.X, (int)p.Y);
   }
}
```

Code 1. Hands and Pointer Handler.

The di Piazza in Piazza application and its interaction system were continuously used by the visitors. Some problems occurred when the users wore very large clothes or large bags or when they tried to interact with the application keeping objects in their hands (e.g. smartphones, books). In those cases, the Kinect sensor detected the users, but it experienced problems in skeleton reconstruction, in particular in joints identification and application. Hence, the interaction system produces some delays and imprecisions. A part these few specific cases, the system generally responded in a good way. Figure 4 and 5 show pictures taken during such an exhibition.

The exhibition in the Museum of the City of Rimini took place in June 2015 for three weeks; in this case a projection was used, hence the configuration set up was the one shown in Figure 2. A room of the museum was exclusively devoted to host di Piazza in Piazza. Data about the average affluence are not yet available. Some more exhibitions in the University of Bologna campuses will be hosted in Cesena (Malatestiana Library, the first European civic library [8]) and in Ravenna (Palazzo Rasponi Dalle Teste) next fall 2015: we are in the process of setting up such events.

Figure 4 - A user enjoys di Piazza in Piazza ("Women and Food" scene).

Figure 5 - Schoolchildren are playing with di Piazza in Piazza in Salaborsa.

VI. Conclusion and Future Work

This paper presents the design issues and the implementation of a hyper-multimedia system that let the users hands-free interact, called di Piazza in Piazza, designed to be a part of cultural performance exhibitions. The paper describes the most interesting and challenging aspects related to the implementation and the putting in practice of the interaction system, which is based on the Kinect sensors.

Di Piazza in Piazza has been yet exposed in two cities (Bologna and Rimini) and the paper reports about such experiences. The project has been selected to be exposed also in Milan, at the EXPO2015, representing one of the Emilia Romagna Region partners of the Universal Exposition. Currently, we are preparing next two exhibitions (in Ravenna and in Cesena) during which we are going to gather data about users' satisfaction, by conducting usability tests with users (through cognitive walkthrough and questionnaires). This will provide more details about the users' experience and involvement in hand-free interacting with, manipulating and even hacking digital contents.

Acknowledgment

The authors want to thank all the colleagues involved in the project. A special thank goes to Dario Braga, Cristina Bragaglia, Maria Giuseppina Muzzarelli, Nelda Parisini, Nicolò Tedeschi for their precious commitment in the di Piazza in Piazza project.

References

[1] M. Roccetti, G. Marfia, and M. Zanichelli, "The art and craft of making the Tortellino: playing with a digital gesture recognizer for preparing pasta culinary recipes," in *Computers in Entertainment*, 2010, Vol. 8(4), pp. 28-47.

[2] M. Roccetti, G. Marfia, and A. Semeraro, "Playing into the wild: A gesture-based interface for gaming in public spaces," in Journal of Visual Communication and Image Representation, 2012, Vol. 23(3), pp. 426-440.

[3] F. Bevilacqua, B. Zamborlin, A. Sypniewski, N. Schnell, F. Guédy, and N. Rasamimanana, "Continuous realtime gesture following and recognition," in *Gesture in embodied communication and human-computer interaction*, Eds. Springer Berlin Heidelberg, 2010, pp. 73-84.

[4] P. Salomoni, L.A. Muratori, S. Mirri, and F. Pozzi, "Hacking Nintendo Wii to paint virtual graffiti," in Proc. Int. IEEE Conf. on Ultra Modern Telecommunications & Workshops, 2009.

[5] Microsoft, "Kinect for Windows," https://www.microsoft.com/en-us/kinectforwindows/ [Retrieved: July 2015].

[6] Michael Nebeling, David Ott, and Moira C. Norrie, "Kinect analysis: a system for recording, analysing and sharing multimodal interaction elicitation studies," in Proc. of 7th ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS '15), 2015, pp. 142-151.

[7] A. Di Iorio, A.A. Feliziani, S. Mirri, P. Salomoni, and F. Vitali, "Automatically producing accessible learning objects," in Journal of Educational Technology & Society, 2006, Vol. 9(4), pp. 3-16.

[8] UNESCO, "Memory of the World Register: The Malatesta Novello Library," http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/nomination_forms/italy_malatesta_novello_library.pdf [Retrieved: July 2015].