Public Systems Supporting Noninstrumented Body-Based Interaction

Dimitris Grammenos, Giannis Drossis and Xenophon Zabulis

Abstract Body-based interaction constitutes a very intuitive way for humans to communicate with their environment but also among themselves. Nowadays, various technological solutions allow for fast and robust, noninstrumented body tracking at various levels of granularity and sophistication. This chapter studies three distinct cases showcasing different representative approaches of employing body-based interaction for the creation of public systems, in two application domains: culture and marketing. The first case is a room-sized exhibit at an archeological museum, where multiple visitors concurrently interact with a large wall projection through their position in space, as well as through the path they follow. The second example is an "advergame" used as a means of enhancing the outdoor advertising campaign of a food company. In this case, players interact with the wall-projected game world through a virtual, two-dimensional shadow of their body. Finally, the third case presents a public system for exploring timelines in both two and three dimensions that supports detailed body tracking in combination with single-hand, two-hands, and leg gestures. Design considerations are provided for each case, including related benefits and shortcomings. Additionally, findings stemming from user-based evaluations and field observations on the actual use of these systems are presented, along with pointers to potential improvements and upcoming challenges.

D. Grammenos (✉) · G. Drossis · X. Zabulis
Foundation for Research and Technology-Hellas (FORTH), Institute of Computer Science, Heraklion, Greece
e-mail: gramenos@ics.forth.gr

G. Drossis
e-mail: drossis@ics.forth.gr

X. Zabulis
e-mail: zabulis@ics.forth.gr

1 Introduction

Human beings employ their bodies, implicitly or explicitly, to interact with their environment but also among themselves. At present, there are various technological solutions available that allow for fast and robust, noninstrumented body tracking at various levels of granularity and sophistication. The term "noninstrumented" denotes that users do not have to carry any object pinpointing their location or the position and pose of their head, arms, or legs. Thus, more natural interactions can be supported also through very simple usage "logistics," as there are no physical items that users have to carry or fasten on themselves—a fact which in the case of public systems can be crucial for their success.

Up to now, three of the most popular approaches for taking advantage of noninstrumented body-tracking technology when developing systems for public use include:

1. *location tracking*. Only the position of the user's body in Cartesian space is taken into account, much like the location of the mouse cursor on a computer screen (e.g., Zabulis et al. 2010).
2. *body-shape tracking*. A two-dimensional (*typically, body-part "agnostic"*) image of the user's body is projected on an interface or a virtual world and used to interact with it through collisions (i.e., overlaps) with virtual entities (e.g., buttons, characters, moving objects) (e.g., Grammenos et al. 2012).
3. *skeleton tracking*. The location and orientation of the user's individual body parts, as well as predefined gestures, are identified and mapped into a virtual world (e.g., Drossis et al. 2013).

Each of the aforementioned approaches comes with its own advantages and drawbacks, thus rendering itself appropriate for a specific range of applications and contexts of use. This chapter provides three case studies—one for each distinct approach—offering related design considerations along with findings stemming from user-based evaluations and field observations on their actual use in practice.

2 Background and Related Work

The notion of *body-based interaction* (though at a more rudimentary level) has been around since Myron Krueger's artificial reality work experiments in the 1980s (Krueger et al. 1985). In the beginning of our century, it has been popularized by Sony's EyeToy and more recently revolutionized by Microsoft's Kinect. One of the earliest application examples of noninstrumented body tracking is the

KidsRoom (Bobick et al. 1999), an interactive playspace simulating a children's bedroom where young children are guided through an adventure story. More recently, Laakso and Laakso (2006) developed a multiplayer game system using one top-view camera where player motion is mapped to digital character 2D motion. Another very popular contemporary example are interactive floors—physical sensor-based, like Magic Carpet (Paradiso 1997) or vision-based, e.g., iGameFloor (Grønbæk et al. 2007)—which are mainly being used for playing games. In the domain of museum applications, Kortbek and Grønbæk (2008) explored three different ways for supporting location-based interaction: (a) a coarse-grained passive infrared sensor; (b) pressure sensors embedded in the floor and a small staircase; and (c) camera tracking. In "Immersive Cinema" (Sparacino 2004), one ceiling-mounted camera is used to track a user's position on a floor segmented in five areas. A different, but quite interesting approach was followed by Robertson et al. (2006) in Bystander. They employed a ceiling-mounted IR camera to track users' position and motion, which are subsequently combined into flocking behavior used to drive the browsing of collections of photographs and texts.

The pose of a user's body can also be tracked and used for selective interaction with the environment (Jaimes and Sebe 2007). For instance, Papadopoulos et al. (2012) use defined body poses recognition in order to allow navigation in 3D environments. Additionally, gestural interaction is widely studied in the literature, but also one of the most popular approaches due to its intuitiveness, as gesture constitute a typical way of interaction among humans and their environment (Nickel and Stiefelhagen 2003; Sangsuriyachot et al. 2011; Yoo et al. 2010). Furthermore, hand gestures can be used to augment systems and allow supplementary interactions (Hilliges et al. 2009) when combined with other means of interaction, such as single- or multitouch. Apart from single-hand gestures both users' hands may be used in combination to enrich the set of available gestures (Fikkert et al. 2010). Finally, the use of feet gesturing (Sangsuriyachot et al. 2011; Valkov et al. 2010) and foot tapping (Crossan et al. 2010; Ronkainen et al. 2007) are occasionally conceded as supplementary helpful interaction tools.

3 Noninstrumented Body Tracking in Practice: The Case Studies

This section presents three cases of different representative approaches to body-based interaction for the creation of public systems, in two application domains: culture and marketing. The first case, Macrographia (Zabulis et al. 2010), is a room-sized exhibit at an archeological museum, where multiple visitors concurrently interact with a large wall projection through their position in space, as well as through the path they follow. The second example, Paximadaki (Grammenos et al. 2012), is an "advergame" used as a means of enhancing the outdoor advertising campaign

of a food company. In this case, players interact with the wall-projected game world through a virtual, two-dimensional, shadow of their body. Finally, the third case, TimeViewer (Drossis et al. 2013), presents a public system for exploring timelines in both two and three dimensions that supports detailed body tracking in combination with single-hand, two-hands, and leg gestures.

3.1 Case A. Macrographia: Multiuser Location-Based Interaction with a Room-Sized Exhibit

In 2010, the Institute of Computer Science of the Foundation for Research and Technology-Hellas (ICS-FORTH) and the Archaeological Museum of Thessaloniki (AMTh) collaborated toward the creation of a special exhibition of prototypical interactive systems with subjects drawn from ancient Macedonia, named "Macedonia from fragments to pixels".[1] The exhibition (Zabulis et al. 2011) comprises seven interactive systems based on the research outcomes of ICS-FORTH's Ambient Intelligence Programme. The digital content of the systems includes objects from the Museum's permanent collection and from ancient Macedonia in general. The largest exhibit is *Macrographia* (Zabulis et al. 2010; Zabulis et al. 2012) a system that presents very large images, which visitors can explore by walking around in a room (Fig. 1). The images are projected on a screen and are analyzed part-by-part depending on the location of each visitor in the room. Macrographia presents the "Wall-painting of the Royal Hunt" from the tomb of Philip II at Vergina, ancient Aigai, the largest ancient Greek painting that has been found to date, its length exceeding 5.5 m. Widely admired as a rare masterpiece of ancient Greek art, the painting shows ten hunters chasing five different animals in a complex landscape.

The exhibit includes a camera network that observes multiple humans in front of a very large display. This network enables the observation of visitors from multiple views (see Fig. 2). The acquired views are used to volumetrically reconstruct and track the humans robustly and in real time using the method of (Argyros and Lourakis 2004), even in crowded scenes and challenging human configurations. The system includes one computer that acquires the corresponding images, processes them, and extracts a spatial representation of the persons in the room. Given the frequent and accurate monitoring of humans in space and time, a dynamic and personalized textual/graphical annotation of the display can be achieved based on the location and the walk-through trajectory of each visitor.

The digital representation of the wall-painting is conceptually separated in five zones perpendicular to the display, based on a semantic interpretation of the themes that appear in it. The room is also conceptually separated in four rows

---

[1] http://www.makedonopixels.org

Fig. 1 Macrographia: a room-sized interactive exhibit at the Archaeological Museum of Thessaloniki with more than 150,000 visitors up to now

parallel to the display, which correspond to different distances of observation. To prevent from continuous alternation in the case of visitors lingering across the boundary of a cell of the grid, the cell size is assumed magnified by 10 % comprising a grid of partially overlapping slots.

At the room entrance, signs guide English-speaking visitors to enter the room by moving rightwards and Greek-speaking visitors to enter the room by moving leftwards. The corresponding textual annotation for each distance is presented at the bottom of the screen (see Fig. 3). Within the context of a zone, the presented content is varied graphically and conceptually according to the distance of observation. When a visitor enters a zone, the presented content matches the viewing distance. The visitor has the capability to explore the corresponding theme by stepping back to get a more abstract view or step closer and focus on the details of the exhibit. When idle and upon visitor entrance, the system presents the wall-painting in its current state. As one or more visitors approach the display, graphical outlines are superimposed to the corresponding region(s) of the display reviving the deteriorated forms. In the next row, the system presents a fully restored version of the painting. In the closest row, the restored version is grayed out, and a specific detail is highlighted, using a combination of color and animation. When multiple visitors stand in the same zone, the person closest to the display determines the content of the presented textual annotation. When this person leaves, the next in line (if any) becomes the closest one to the display. By tracking visitors and assigning a unique identifier to each one, the system also retains attributes for each
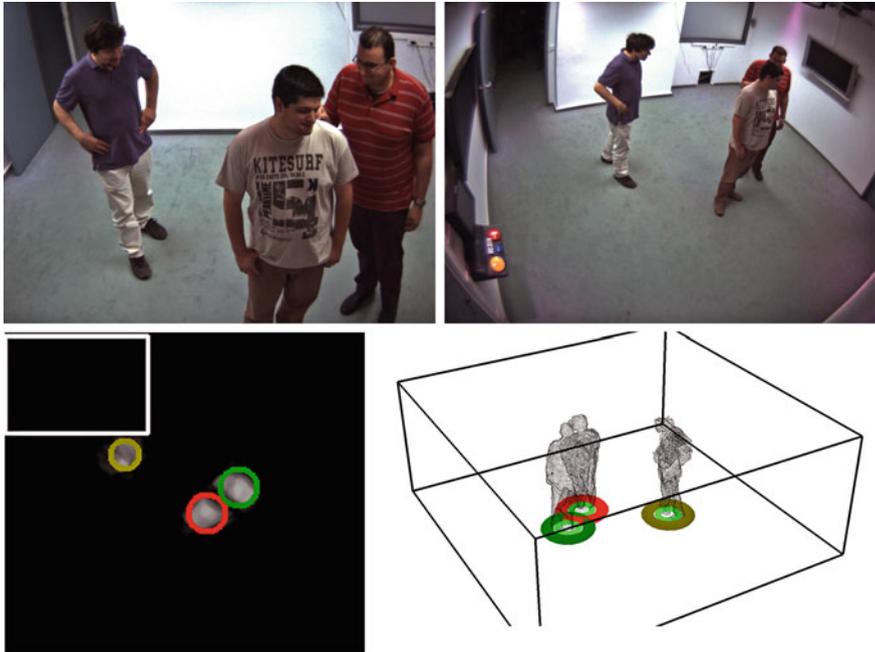
Fig. 2 Person localization and tracking. (*Top*) original images from two out of eight views, (*below*) person segmentation and 3D reconstruction. Person tracking results are rendered as circles, superimposed on the ground plane of the 3D reconstruction. Circle colors correspond to track ids. Tracking retains the correct id for all persons although it is often that visitors may come in contact or occlude each other to the cameras

Fig. 3 Sample of the projected image as presented when three distinct persons are located in the room at different positions. The leftmost and rightmost parts are presented in their original state, as no one is standing in front of them. The second from the left part presents a sketch and a descriptive title, the next one a fully reconstructed view and a detailed description and the next part highlights a detail of the image, also offering related information

of them. Using an attribute for the language, which is set upon visitor entrance, the textual components of the presented content are provided in the language selected by each visitor. Furthermore, using collected data regarding the user's trajectory in space, as well as the time spent on each slot, the system has the ability to present additional content (this feature is currently used experimentally at an installation of the exhibit at the premises of ICS-FORTH).

As a means of providing real-time visual feedback, except the highly visible changes that happen to the image, a more subtle—but continuous—cue is also provided. Underneath the image, to the left of the information box, there is a small triangle that directly maps to a user's distance from the wall (see bottom part of Fig. 3); i.e., when a user moves toward the wall, the triangle moves downwards and vice versa. The area upon which this arrow moves is separated in three parts, which correspond to the available three levels of information. The currently selected level is drawn in green color, while the rest in red.

During the development of the system, ethnographic field methods (Blomberg et al. 2003) were employed, using a combination of the "observer participant" and "participant observer" approach. During a 6-months period, more than 200 persons have participated. Using the final version of the system, summative evaluation sessions took place both at the premises of ICS-FORTH, as well as at the Archaeological Museum of Thessaloniki (Zabulis et al. 2012). For the purposes of evaluating the experience of the users with the system, a 13-item attitude Likert scale questionnaire was created which was based on Brook's System Usability Scale (SUS) questionnaire (Brooke 1996). Evaluation at ICS-FORTH took place in a laboratory room that was set-up in a way to resemble as close as possible the actual museum exhibit where this application is currently being housed. Twenty-two volunteers participated (13 male, 9 female). The average age of the participants was 31.7 years old, the youngest being 18 and the oldest 41 years old. No specific instructions were given to the participants as to how the Macrographia system actually worked or displayed the information in order to examine if they were able to understand how the system actually worked and how to retrieve the information that corresponded to each section. At the Archaeological Museum of Thessaloniki 22 questionnaires were filled out by 15 visitors and 7 guards. Of those, 11 were males and 11 were females. Their average age was 34.6 years, the youngest being 21 and the oldest 56 years old.

The overall usability of the system was rated high in both studies: 82.8 % in-house and 80.8 % at the museum. The questionnaire results were also supported by data collected by analyzing the recorded sessions or interviewing the participants for the in-house evaluation, and by hand-written comments of the museum visitors. Overall, the comments made by the all participants were highly positive. Most of them were impressed with the system's ability to track accurately their position in the room and display the information in the language that was chosen. Even though very little instruction was given to them before they entered the room, they all managed to understand that their movement was tracked and that the information changed according to their distance from the screen. As a result, all users were able to read all the information that was presented in each section of the Macrographia. They also offered some suggestions on how to improve it. For
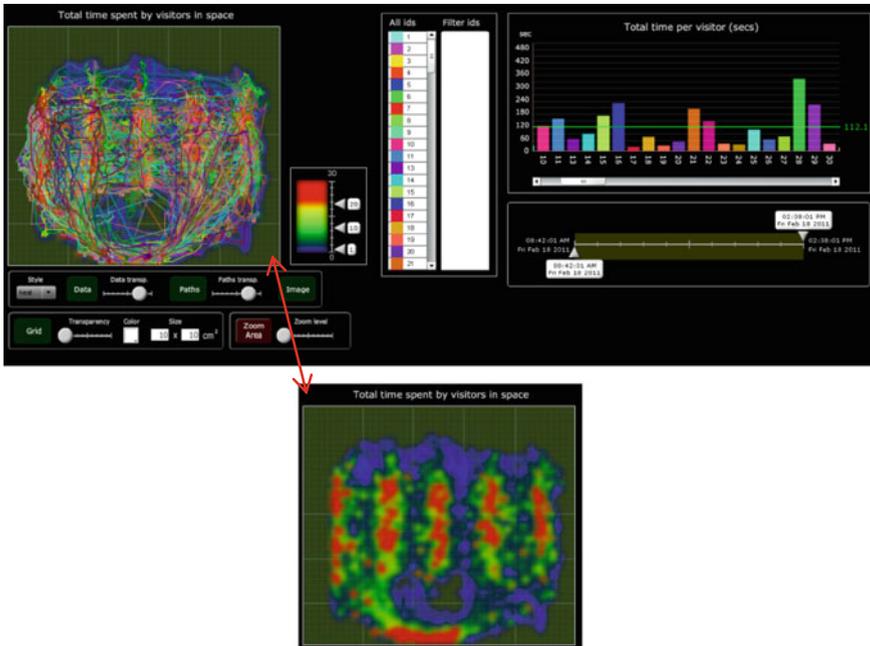
Fig. 4 Interactive visualization component for analyzing visitor data *top* showing visitor paths; *bottom* heat map (partial view of the interface)

example, a few users suggested that it would be better if the text font size changes from larger to smaller the user approaches the display.

An interactive visualization component which presents information about the exhibit's visitors was also developed (Stamatakis et al. 2011). The main part of the visualization component's user interface comprises a top-down view of the room and its interactive slots. On this view, user paths are presented (using a different color for each distinct user—Fig. 4, top). Additionally, accumulated user time spent on a specific area is illustrated in the form of a heat map (Fig. 4, bottom). Quantitative information about each point of the heat map can be seen by positioning the mouse pointer over it. Additionally, several view parameters can be modified, for example, changing the transparency level of the paths or the heat map, hiding the background image, overlaying a grid, and changing the coloring scheme of the heat map. Furthermore, zoom is supported in any part of the view at different levels. The interface also presents the total time spent by each visitor in the room, the average time for all visitors and provides the means to filter out data by selecting a subset of IDs. Finally, there is an interactive control for setting a time period of interest that can be used for filtering data according to the time of their creation. The visualization component can work with both real time and stored data. Up to now, the analysis of the collected data from the museum installation has not revealed any prevalent patterns regarding the trajectory of

visitors in the room. By examining the generated heat map views, it becomes evident that although the interactive slots are not marked on the floor, there is a significant concentration of visitors' "footprints" near their centers. Also, as expected, the furthest slots, where the presented information is minimal (just a title), have a much smaller aggregate visit time than the rest.

3.2 Case B. Paximadaki: An Advergame Public Installation Controlled Through Virtual Body Shadows

Paximadaki (Grammenos et al. 2012) is an advergame installation targeted to promoting, in exhibition spaces and key points of sale, the products of a food company producing various types of traditional Cretan rusks. The game, entitled "Paximadaki" (*small rusk* in Greek) is a Kinect-based PC exergame (Mueller et al. 2003) projected against a large surface, involving physical activity as a means of interaction. The main reasons for selecting Kinect were that it allows for noninstrumented game control through natural movements, which also afford higher levels of engagement and social behavior (Lindley et al. 2008), performs well under various environmental conditions, and comes at a very low cost. In this respect, it was decided to just use the depth camera's image in order to render a virtual shadow of the players, instead of tracking body skeletons. The rationale is two-fold. On the one hand, it is easier for people, especially "non game-players", to identify with their shadow rather than with an avatar, thus achieving a higher level of control and immersion (Sweetser and Wyeth 2005). On the other hand, this approach allows for maximum flexibility regarding the number, posture, and size of players, as well instantly joining and leaving the game, thus maximizing the opportunities for social interaction (Sweetser and Wyeth 2005). The downside is that people with larger body sizes have an advantage, and there also exists the possibility of accidental "intrusions" in the play area.

The gameplay[2] is simple, straightforward, and has very clear goals (Sweetser and Wyeth 2005). Players perceive their bodies as shadows projected on a brick wall (Figs. 5 and 6). On the startup screen there are buttons for setting the game's difficulty and starting a one- or two-player game. Players can select these buttons by overlapping any part of their shadow on them for a specific period of time (i.e., dwell), which is indicated by a countdown that appears right upon the button. Depending on the players' number, there may be one or two baskets at the two bottom sides of the wall (Fig. 6). A "rainfall" of rusks starts. Players must use their shadows to put the rusks into their basket. Rusks that fall on the floor are broken into pieces. The game ends when a certain number of rusks have fallen. In order to deliberately create a "memorable moment" (Jenkins 2002), when the end of the game approaches, a huge amount of rusks suddenly start to fall. To notify

---

[2] Videos of indicative play sessions can be found at: http://www.youtube.com/user/icsforthami.

Fig. 5 Paximadaki: an advergame installation promoting the brand and products of a food company, installed at five different public events and played by more than 1,500 people with ages ranging from 2 to 76 years old

players about this event, a bleating goat appears, providing a humorous note. Additionally, the music shifts to a faster tune. During gameplay, at moments that are likely to provoke interesting players' poses the game automatically takes photos of them. Additionally, when the game ends, a photo countdown ("smile moment") appears allowing players to pose. The photos are presented on screen when the game finishes and can be sent to the players via e-mail. Each game session has a unique serial number that appears on the bottom right corner of the screen. When the game finishes the players can write down (on a tablet or a paper pad) their e-mail address on a list next to the game's serial number in order to receive the photos.

Up to now, the game (with some variations) has been installed with remarkable success in five different public events in key locations in Athens, Greece, where it has been played by more than 1,500 people of ages ranging from 2 to 76 years old. When a complete version of the game was available, more than 50 h of playtesting along with the employment of observational usability assessment methods took place in order to fine-tune the gameplay and also debug the game, in a realistic installation at the premises of ICS-FORTH, with more than 30 players of both

Fig. 6 Screenshot of indicative gameplay showing a player's virtual shadow against the wall

genders, with ages ranging from 4 to 52 years old. Detailed usability evaluation results can be found in (Grammenos et al. 2012).

Subsequently, the game was installed in an exhibition of traditional food products, in the central metro station of Athens. In the course of three days, the game was played by 173 players (127 female, 46 male). Their age ranged from 3 to 76 years. All players stated that they liked the game and that they enjoyed playing it. More than 100 people used words like "fantastic," "great," "very good," or another synonym. As a means of collecting qualitative data about the game and its impact, a 20-item online questionnaire was created. 25 people answered it (18 female, 7 male) with ages ranging from 18 to 60. Among the things that respondents mentioned were that they liked: the responsiveness of the game; the (easy) way it was controlled; its high quality; its originality; the fun they got out of it; the fact that their whole family could play it; the music.

The game was then installed at the Zappeion Exhibition Hall, in Athens. In three days, the game was played by 337 players (195 female, 142 male) from 2 to 75 years (played against his 70-year-old mate). In 20 cases, the game was concurrently played by three or more players. The maximum number of concurrent players was 6. When more than two players played the game, their positions in space were usually dictated by their age and personal relationships. For example, friends or children of the same age would stand next to each other (see Fig. 5), while parents would mostly stand behind their children trying to discretely help them. When there were up to four players, they would usually form 2 teams. When there were more, people would not pay much attention to the score or in forming teams, focusing mainly on having fun through the physical aspects of play.
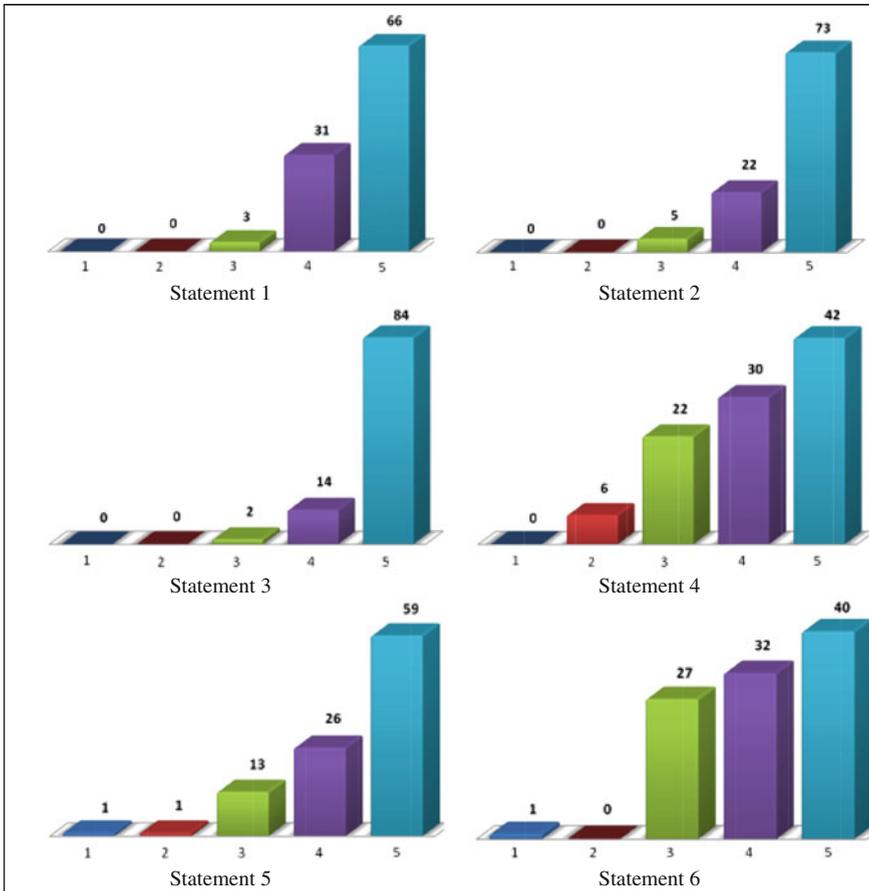
Fig. 7 On-the-spot evaluation questionnaire scores by 100 respondents (second installation of the game—Grammenos et al. 2012)

In 55 cases children, played along with their (grand)parents—sometimes both of them were adults. This time, a shorter questionnaire was employed and a total of 100 people responded (61 female and 39 male). The questionnaire included a few demographic-related questions and six statements.

1. I liked it.
2. It was easy to learn how to play.
3. It was fun.
4. It helped me familiarize with the brand *<brand name>* and its products.
5. It contributed toward creating a positive image about *<brand name>* and its products.
6. It positively affected me toward purchasing products of *<brand name>* .

Questionnaire scores are illustrated in Fig. 7. Interestingly, there were 28 respondents who scored 5/5 (i.e., best score) all statements. In short, players believed that the game was indisputably very fun and very easy to learn how to play. The related statements in all questionnaires were the ones with the highest average ratings and the least standard deviations. Additionally, as was deducted from both the analysis of the questionnaires (Grammenos et al. 2012) and on the spot questioning of the players, the game offered a highly entertaining experience while, at the same time, positively contributing to the marketing of the company and its products, reinforcing previous research findings.

The third installation of the game was in Athens Plaython, a city play event. The most interesting fact in this case, was that out of the 150 game sessions that took place in the course of 2 days 127 were played by more than two players, with an average of four players and a maximum of 12 players! In the latter case, players created a compact mass with several protruding hands and were mostly interested in just hitting the falling objects (e.g., much like playing volleyball on the beach), irrespectively of whether these would land in a basket or not. It should be noted that, due to appropriate design considerations, even in such extreme cases neither game performance nor play time were negatively affected. No formal evaluation was conducted in this installation.

Subsequently, there have been two more installations of the game at food exhibitions, while since July 2010, a new version of the game, targeted to familiarize visitors of the island of Crete with local products, is permanently installed at the arrivals hall of Heraklion international airport "Nikos Kazantzakis" as part of *Creative Crete,*[3] a project realized by the Institute of Computer Science of FORTH under the initiative of the Region of Crete.

3.3 Case C. TimeViewer: A System for Exploring Timelines in Two and Three Dimensions Through Hand and Body Gestures

TimeViewer (Drossis et al. 2013) is a system allowing the modeling, storing, visualization, and multimodal interaction with timelines (Fig. 8). For example, up to now it has been used to present historical information (e.g., the history of computers, the history of a city), the life trajectory of famous personalities, or even time threads comprising important events, work, and achievements of various entities. A principal characteristic of TimeViewer is the provision of an immersive display of temporal information supplementary to a view that favors the provision of an overview. In this direction, TimeViewer supports two distinct co-existing modes (i.e., views) for visualizing information, respectively (Fig. 9): the 'Tunnel' and 'Classic2D'. Users can freely and seamlessly switch between the two modes at

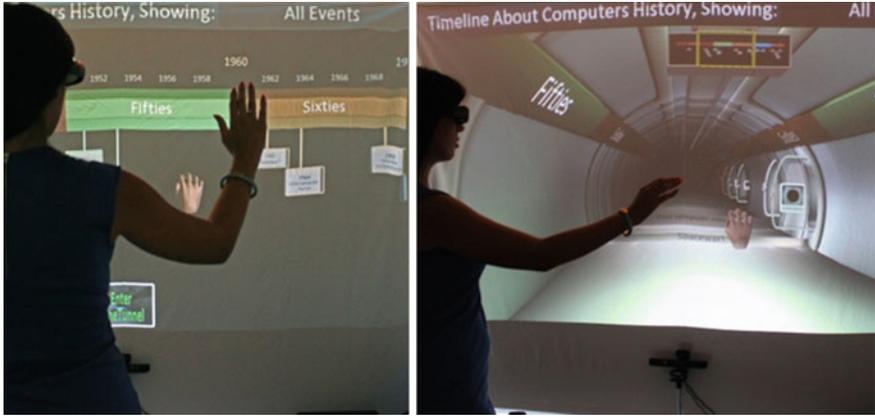---

[3] https://www.facebook.com/creativecrete

Fig. 8 TimeViewer: A system allowing the modeling, storing, visualization, and noninstrumented body-based kinesthetic interaction with timelines. Up to now, it has been evaluated inhouse with 16 representative users

Fig. 9 The Classic2D (*left*) and Tunnel (*right*) visualization modes

any time. Classic2D mode is used to provide an overview of the available information in a manner familiar to the user, easy to understand, and convenient for navigation (Fig. 9, left). The Tunnel mode (Fig. 9, right) is targeted to supporting the task of sequential, exhaustive, exploration of the presented information, allowing the user to physically experience the time dimension as well as the unambiguous display of each event. The time dimension is visualized as a "tunnel" along which all information is integrated. The tunnel walls serve a dual role: they considerably constrain navigation in 3D space, also offering rich orientation cues, and can host contextual information, such as event-related items. The user's perception of the time currently inspected is enhanced by the presence of a minimap in the form of a railway wagon, which is placed at the tunnel's roof, covering the whole timeline extent. The minimap rendered is the Classic2D representation along with a frame that provides live feedback regarding the equivalent position in the other view of TimeViewer, while also acting as a gateway to Classic2D view.

TimeViewer supports a variety of multimodal interaction techniques, even in combination, so as to offer natural interaction in a wide range of hardware set-ups. Apart from common desktop-based interaction techniques (i.e., point and click or multitouch gestures), TimeViewer supports noninstrumented kinesthetic interaction using Microsoft's Kinect, targeted to navigation and manipulation of 3D virtual environments. Mimicking a person's actions in the real world, users can employ their feet to move and their hands to reach out to virtual elements. More specifically.

- *User Position.* In Classic2D mode the user can stand off-center, toward either side of the display, in order to scroll the projected view toward the respective direction. Furthermore, the system interprets the distance of the user from the display as a zoom mechanism. In the Tunnel mode, the system maps the actual position to the place from which the user looks in the tunnel, creating the illusion of "being there".
- *Virtual Hand.* A cursor, in the form of a three-dimensional virtual hand, appears when the user extends his or her hand toward the display, following and mapping the user's real hand. Items' selection is accomplished by keeping the hand cursor over them for a short period of time.
- *Hand Gestures.* In several cases, when the usage of the virtual hand is not optimal or appropriate, TimeViewer additionally supports single- and dual-hand gestures. The simplest single-hand gesture supported is the movement of a hand for scrolling lists of multimedia elements. Another example involves raising a hand while turning the torso in the same direction, to rotate the projected view around its vertical axis. Dual-hand gestures include pushing and pulling information objects in order to select/deselect them (i.e., bring them closer or send them away), moving between events in Tunnel mode, or examining detailed information about an event in Classic2D mode.
- *Leg Gestures.* TimeViewer adopts the notion of stepping: when a user steps (with one leg, holding the other still) right, left, forward, and backward, navigation is achieved in the corresponding direction. In Classic2D mode, stepping forward or backward results in increasing or decreasing the scale of the timeline respectively, while stepping left and right results in exploring the timeline's display at the respective direction. In Tunnel mode, stepping forward or backward allows navigation along the tunnel, while stepping at the side results in displaying the tunnel as if the user is looking from the equivalent side.

To assess TimeViewer in terms of usability and user experience but also to compare the various supported interaction methods, an evaluation study (Drossis et al. 2013) was conducted with 16 participants (7 female and 9 male) the age of whom varied from 20 to 40 years old. Participants were instructed to express their thoughts aloud (Think Aloud) throughout the completion of each task. In addition, at the end of the evaluation they were requested to fill out a Likert type of questionnaire.

The enthusiasm of the users was apparent during their interaction with TimeViewer when using body tracking, as only one user stated his preference to interact with the system using more traditional modes. Users were captivated by the remote handling of the system achieved without the use of any wearable component.

Hand-related interaction was regarded as very natural, as indicated by the comments during the evaluation and the completed questionnaires. All users found the gestures representative of their functions and the only concern of a small percentage of them (4/16) was that they might result in fatigue after prolonged interaction. Gestures were also considered easy to learn, as all users were generally able to accomplish them on their own after being shown by the evaluator and largely preferred them to the use of the virtual hand cursor, the manipulation of which also came easily to the users. A common remark involved the cursor's (in)stability, which created some accuracy problems with small-sized items. Furthermore, some of the users stated that they would prefer shorter dwell times.

Leg gesturing was almost unanimously accepted. Only one user had trouble with it, as Kinect failed to successfully recognize the exact placement of his legs due to the material of his trousers. Especially nonexpert users were more in favor of leg gestures as they felt more comfortable with handling the system more naturally and less unambiguously. Participants did not face any problem understanding the related conceptual model and found it efficient and fascinating, a fact which is clearly indicated in the corresponding questionnaire answers.

In conclusion, although kinesthetic interaction was considered as fractionally less responsive and more tiring to use, both hand and leg gesturing were deemed as preferable to touch interaction. Despite the higher user error rates in comparison to interaction through touch, users were almost unanimously (15 out of 16) in favor of kinesthetic interaction. This is mainly attributed to the fact that the fun factor was considered as more important than higher accuracy.

4 Discussion

If we had to use just a couple of words to describe each of the aforementioned approaches to body tracking, based on the users' reaction to the respective application then these words would be:

- Macrographia (*location tracking*): Surprising and Impressive
- Paximadaki (*body-shape tracking*): Fun and Intuitive
- TimeViewer (*skeleton tracking*): Versatile and Powerful.

Macrographia made a memorable impression to its users. The moment people walked in the room (especially when there was no one already in) and saw the "wall" responding to their movement, their facial and verbal expressions articulated the feeling of positive surprise and excitement. This phase was followed by a "curiosity" period during which they tried to discover and explore all the available pieces and modes of information. Then again, Paximadaki had a more "cozy" effect. The fact that people saw their own shadow on the virtual wall, as well as their face in the photographs at the end of the game, created an affinity toward the game, which was subsequently directed toward the people running the installation (see Grammenos

Table 1 Advantages and limitations of the 3 presented approaches to employing body tracking

Type of tracking	Advantages	Limitations
Location	• Robust • Multiuser (number depending on available space) • Can work well with very large displays • Can work in large spaces • Requires limited user mobility	• Small "interaction vocabulary" • "Midas touch" • Can support coarse-grained interaction • Limited number of users • Complex to implement • Requires (considerable) space • Limited potential applications • Requires ability to move in space • Indirect user control • Higher cost
Body-shape	• Very robust • Very intuitive • User perceives own image • Multiuser (virtually unlimited number of concurrent users) • Very flexible, works for highly diverse user characteristics • Very quick response • Easy to implement • Low cost	• Small "interaction vocabulary" • "Midas touch" • Limited potential applications (mostly games) • Requires ability for (gross) body movement • Prone to interferences (from objects or other people in range)
Skeleton	• Rich interaction vocabulary • Multiuser (limited number of concurrent users) • Highly versatile, essentially unlimited application types supported • Can provide very natural interaction • Can support tasks requiring high accuracy • Low cost	• Lless robust • "Midas touch" (in some cases) • More 'rigid' (more strict user motion), especially when gestures are used • May not work (properly) for some body types, clothes, accessories • Requires ability for fine body movement and control • More complex to implement • Harder to learn and use • (Can be) less intuitive, may require training

et al. 2012). Furthermore, the selected interaction approach was very intuitive indeed. In all installations 30–40 % of the players had never played any type of video game before and had no problem mastering game control after the first few seconds. Probably the best example illustrating this fact was 2-year-old Maria who, after playing the game and being all excited about it, would wander around explaining to anyone standing nearby how to play: "Big basket, rusks put in—many many rusks, yeaah!" Finally, TimeViewer provides much richer and fine-grained interaction supporting a large "interaction vocabulary" which allows designers to apply the most appropriate interaction technique to each task, and, often, users to select among redundant techniques according to their preferences. Additionally, maybe also due to the novelty of kinesthetic interaction, users showed significant preference and liking toward it, despite its potential fatigue and accuracy issues.

Table 1 provides an overview of the individual advantages and limitations of each approach.

In principle, all three approaches can accommodate multiple users, but at different scale. Skeleton tracking, currently works best for a very small number of users. Body shape tracking has no technical limitations (we have seen up to 12 players at Paximadaki) but in practice, there is a certain point where a crowded shadow image "overflows" the game terrain, rendering interaction meaningless. Location tracking can scale up, but at the cost of scaling up the interaction space, as well as of the technical equipment used.

The first two approaches have a very limited "interaction vocabulary", i.e., distinct "commands" that users can issue through them. Skeleton tracking, especially when hand and leg gestures are used, can have a very rich and context-related "interaction vocabulary" but at the cost of ease of use, intuitiveness, and requirement for user training.

All three approaches have considerable user mobility requirements with skeleton tracking being the most demanding one, both in terms of gross motor skills (i.e., movement of the arms, legs, feet, or entire body) and endurance (can become quite tiring after a while).

A common problem of all approaches is what is usually referred to as "Midas touch," or, in other words, identifying when users want to perform an action to an interactive object that is currently under their "reach" (e.g., a slot that corresponds to their location in Macrographia, a button that overlaps with their virtual shadow in Paximadaki, or an object located underneath the hand cursor in TimeViewer). In all cases, a common solution is "dwell time," where the respective action is triggered after a predefined time period. Another possible solution that can work in some cases is to require a specific combination of overlaps in time and space which would be hard to happen by accident. For example, in order to start the game, a player may have to sequentially "touch" 3 different objects or even the same object 3 times, but every time it is touched it moves to a different location. Such approaches are appropriate for actions that require confirmation and do not have to be frequently performed. Additionally, if a richer interaction vocabulary can be supported (like in the case of TimeViewer), more elaborate solutions can be provided (e.g., activation gestures).

5 Conclusions

All collected pieces of evidence stemming from the evaluation and observational data of the three case studies converge to the fact that people of all ages and backgrounds indisputably like and enjoy body-based interaction with public systems. On the one hand, it feels more natural and intuitive, while one the other hand it "frees" them from the highly static and rather mundane interaction with desktop and mobile computing platforms. Furthermore, in general, body motion is conceived as more "fun," irrespectively of the actual application's content. For instance, unlike Paximadaki, both the Macrographia and the TimeViewer systems provide "serious" (educational) information, but most users commented that they

were fun to use. Furthermore, in several cases, a considerable number of concurrent users can be supported, thus effectively allowing for group interactions (coordinated or not), which are highly valued (especially among families and friends) and are currently underserved (if at all) by most public systems. In more practical terms, another important aspect of noninstrument body-based interaction with public systems is that users do not come in physical contact with any part of them, thus avoiding dirt, wear and (intended or not) damage of the equipment, contributing to its longevity but also solving several public hygiene issues.

On the downside, body-based interaction also comes with some inevitable weaknesses. First of all, as requirements for body motion and control increase, so do the potential accessibility problems for people with mobility limitations or diversifications, due to circumstantial events (e.g., carrying something), temporary (e.g., a broken leg) or permanent (e.g., paraplegia) disabilities, age (e.g., young children), etc. Then, as the domain is still emerging, there are no widely used and established conventions—a common "interaction vocabulary"—that users are familiar with. As a result, quite often users may not be aware of what they are able to do with a system and some training may be required (even for apparently "intuitive" interactions), and, in some cases, interaction approaches of different systems can be completely inconsistent. To this end, the level of instruction that should be provided in public systems, as well as, how this should be achieved, is a key open issue. Finally, due to the large diversity of human bodies, the items that may be residing on them (e.g., clothes, accessories, devices), as well as of specific environmental conditions, there will always be cases were a particular body-tracking approach may not (properly) work.

A very interesting aspect of noninstrumented body-based interaction is that it inherently supports various forms of formal and informal social interaction. As it takes place in considerable space, a number of actors can be easily involved, in various roles. For example, in the case of location-based interaction (Macrographia) typical "museum-related" social behavior (e.g., politeness—avoid to stand in front of another person) was observed, but also playful activities (e.g., a couple hugging trying to share the same slot), including ad hoc group forming (e.g., a number of strangers standing next to each other in order to reveal all the layers of information of a specific level). In the case of Paximadaki, except the types of social interaction already described in the related section (e.g., cooperation, competition, unstructured group play), another noteworthy behavior was the occasional instantaneous "intrusion" of spectators in the play area, for example to tease a friend by blocking a basket, or extending a helping arm to aid the weaker player. Finally, as the third system presented was targeted for use by a single user, the respective social behavior was limited to people switching roles by going closer or retreating from the screen. Still, the described approach can be (and is currently) extended to support both synergetic and parallel interaction by at least two users.

References

Argyros AA, Lourakis MIA (2004) Real time tracking of multiple skin-colored objects with a possibly moving camera. In: Proceedings of the European conference on computer vision (ECCV'04), vol 3. Springer, Prague, Czech Republic, pp 368-379, 11–14 May 2004

Blomberg J, Giacomi J, Mosher A, Swenton-Wall P (2003) Ethnographic field methods and their relation to design. In: Participatory design: principles and practices. Lawrence Erlbaum Associates, pp 123–155

Bobick AF, Intille S S, Davis JW, Baird F, Pinhanez CS, Campbell LW, Ivanov YA, SchütteA, Wilson A (1999) The KidsRoom: a perceptually-based interactive and immersive story environment. Presence: teleoper. Virtual Environ 8(4):369–393

Brooke J (1996) SUS: a quick and dirty usability scale. Taylor and Francis, London, pp 189–194

Crossan A, Brewster S, Ng A (2010) Foot tapping for mobile interaction. In: Proceedings of BCS '10, British Computer Society Swinton, pp 418–422

Drossis G, Grammenos D, Adami I, Stephanidis C (2013) 3D visualization and multimodal interaction with temporal information using timelines. In: proceedings of INTERACT 2013, lecture notes in computer science, vol 8119. Springer, Heidelberg, pp 214–231

Fikkert W, van der Vet P, van der Veer G, Nijholt A (2010) Gestures for large display control. In: Proceedings of GW'09, LNCS 5934. Springer, Heidelberg, pp 245–256

Grammenos D, Margetis G, Koutlemanis P, Zabulis X (2012) 53.090 Virtual rusks = 510 real smiles using a fun exergame installation for advertising traditional food products. In: Nijholt A, Romão T, Reidsma D (eds) Advances in computer entertainment, LNCS 7624. Springer, Heidelberg, pp 214–229

Grønbæk K, Iversen OS, Kortbek KJ, Nielsen KR, Aagaard L (2007) iGameFloor: a platform for co-located collaborative games. Proceedings of ACE '07, vol 203. ACM, New York, pp 64–71

Hilliges O, Izadi S, Wilson A, Hodges S, Mendoza AG, Butz A (2009) Interactions in the air: adding further depth to interactive tabletop. In: Proceedings of UIST '09. ACM, New York, pp 139–148

Jaimes A, Sebe N (2007) Multimodal human–computer interaction: a survey. Journal computer vision and image understanding archive, vol 108(Issue 1–2). ACM, New York, pp 116–134

Jenkins H (2002) Game design as narrative architecture. In: Harrington P and Frup-Waldrop N (eds) First person. MIT Press, Cambridge, pp 118–130

Kortbek KJ, Grønbæk K (2008) Interactive spatial multimedia for communication of art in the physical museum space. In: Proceeding of MM '08, pp 609–618

Krueger MW, Gionfriddo T, Hinrichsen K (1985) VIDEOPLACE—an artificial reality. In: Proceedings of CHI'85, San Francisco, pp 35–40

Laakso S, Laakso M (2006) Design of a body-driven multiplayer game system. Comp Entertain 4(4):7

Lindley SE, Le Couteur J, Berthouze NL (2008) Stirring up experience through movement in game play: effects on engagement and social behaviour. In: Proceeding of CHI '08, pp 511–514

Mueller F, Agamanolis S, Picard R (2003) Exertion interfaces: sports over a distance for social bonding. In: Proceedings of CHI '03, pp 561–568

Nickel K, Stiefelhagen R (2003) Pointing gesture recognition based on 3D-tracking of face, hands and head orientation. In: Proceeding of ICMI '03. ACM, New York, pp 140–146

Papadopoulos C, Sugarman D, Kaufmant A (2012) NuNav3D: a touch-less, body-driven interface for 3D navigation. In: Proceeding of IEEE VR 2012. IEEE, pp 67–68

Paradiso J, Abler C, Hsiao K, Reynolds M (1997) The magic carpet: physical sensing for immersive environments. In: CHI '97 extended abstracts, pp 277–278

Robertson T, Mansfield T, Loke L (2006) Designing an immersive environment for public use. Proceedings of PDC '06. ACM, New York, pp 31–40

Ronkainen S, Häkkilä J, Kalev S, Colley A, Linjama J (2007) Tap input as an embedded interaction method for mobile devices. In: Proceeding of TEI'07. ACM, New York, pp 263–270

Sangsuriyachot N, Mi H, Sugimoto M (2011) Novel interaction techniques by combining hand and foot gestures on tabletop environments. In: Proceeding of ITS '11. ACM, New York, pp 268–269

Sparacino F (2004) Scenographies of the past and museums of the future: from the wunderkammer to body-driven interactive narrative spaces. In: Proceeding of MM '04, pp 72–79

Stamatakis D, Grammenos D, Magoutis K (2011) Real-time analysis of localization data streams for ambient intelligence environments. In: The proceedings of AmI 11: international joint conference on ambient intelligence, Amsterdam, Springer, Berlin, Heidelberg, pp 92–97, 16–18 Nov 2011

Sweetser P, Wyeth P (2005) GameFlow: a model for evaluating player enjoyment in games. Comp Entertain 3(3):1–24

Valkov D, Steinicke F, Bruder B, Hinrichs K (2010) Traveling in 3D virtual environments with foot gestures and a multitouch enabled world in miniature. In: Proceeding of VRIC 2012. IEEE, pp 171–180

Yoo B, Han J-J,Choi C, Yi K, Suh S, Partk D, Kim C (2010) 3D user interface combining gaze and hand gestures for large-scale display. In: Proceeding of CHI '10. ACM, New York, pp 3709–3714

Zabulis X, Grammenos D, Argyros A, Sifakis M, Stephanidis C (2011) Macedonia: From Fragments to Pixels. ERCIM News Special Theme: ICT Cult Heritage 86:25–26

Zabulis X, Grammenos D, Sarmis T, Tzevanidis K, Argyros AA (2010) Exploration of large-scale museum artifacts through noninstrumented, location-based, multi-user interaction. In: Proceedings of VAST'2010, Palais du Louvre, Paris, France, pp 155-162, 21–24 Sept 2010

Zabulis X, Grammenos D, Sarmis T, Tzevanidis K, Padeleris P, Koutlemanis P, Argyros AA (2012) Multicamera human detection and tracking supporting natural interaction with large scale displays. Mach Vis Appl J 24(2):319–336. Feb 2013