Day and Night at the Museum: Intangible Computer Interfaces for Public Exhibitions

Marco Roccetti, Gustavo Marfia and Cristian Bertuccioli

University of Bologna, Mura Anteo Zamboni 7, 40127 Bologna, Italy

Abstract

Computer technologies have been adopted in many different venues, including public exhibitions and museums, as they can easily support the exchange of natural interactions and provide unimaginable exploration tools of masterpieces and exhibits. This has led many to design and implement a plethora of different technologies for the detection, tracking and action recognition of visitors within a museum. Nonetheless, no single approach has been firmly accepted so far, as it typically suffers from the limitation of adopting separate techniques for detecting, tracking and recognizing the actions of visitors. The contribution of this paper is that of filling this gap: we propose a unifying methodology through which all of the abovementioned services can be handled within a museum. Furthermore, such methodology results being: (a) simple to implement, (b) non-invasive and (c) requiring minimal hardware resources. As significant evidence, we present the experimental results drawn from two relevant implementations: Mercator Atlas Robot exhibited at the Poggi Palace Museum of Bologna and Tortellino X-Perience at the World Expo held in Shanghai. Finally, we indicate how the presented approach can be extended to efficiently support any interaction with several visitors simultaneously.

Keywords: *Digital Technologies for Museums; Intangible Interfaces; Gestural interfaces; Design Interaction; Image Processing; Immersive Environments.*

---

1. Introduction

Digital technologies are slowly, but undauntedly finding their way in all of the domains of interest to human beings. Public exhibitions and museums represent no exception, as a central role is already being played by technology in the related fields of entertainment and education [1-4]. However, unlike other sites, museums and public exhibitions are places that present a few distinguishing characteristics and requirements:

1. The presence of any technology cannot be invasive as museums, especially in old-world countries, occupy historic buildings whose look cannot be changed in any way;
2. The use of HCI technologies is well accepted as long as they reveal their utility in the process of turning exhibitions from containers into live and engaging experiences, without gaining the leading role themselves. In fact, museum professionals often fear the occurrence of Guggenheim Effects, i.e., all those situations where visitors are fascinated more by the *eye-candy* that is presented to them (e.g., any advanced visual effect) than by the masterpieces that they preserve [5, 6].
3. Museums often contain masterpieces and artifacts that are part of itinerant exhibitions, hence any employed technology should be able to meet the requirement of being flexible, i.e., easily removable when an exhibit leaves and rapidly adaptable when a new one is displayed.

Under such premises, we here deal with the problem of how HCI technologies can be effectively designed to play according to the aforementioned rules and requirements, while acting the role of mediators between visitors and exhibits. Such task cannot be accomplished without first crafting the architecture of those required services. A few years of experience in this domain, added to a survey of relevant literature, confirm that three types of services are being extensively sought:

(a) Visitor detection, which is at the base of all those interactive performances that require the recognition of a visitor at a given location (e.g., activate a video reconstruction of the Battle of Waterloo when a visitor touches a virtual button in the Louvre's Napoleon room);

(b) Visitor tracking, that is utilized by those systems that, depending on the path taken to reach a given position, generate a different type of immersive experience (e.g., track the body or the upper limbs of a visitor and produce some type of feedback, sounds coming from gladiators battles at the Colosseum vs lyre music, based on the path that is taken);

(c) Visitor action recognition, employed by all those applications that respond to specific actions (e.g., flip a page of a projected representation of a book when a visitor swings his/her hand).

Now, given the demand for a systematic use of new means of mediation between visitors and exhibits, based on the gestures and the movements that visitors' bodies perform, the technical problem amounts to seeking for a unifying

methodology that is capable of supporting both the detection of the presence of a visitor at a given point and the tracking and interpretation of multiple body points (i.e., head, hands, feet) for real-time action recognition. Furthermore, the sought solution should easily adapt to dynamic scenarios, where the content, as well as the arrangement, of an exhibition changes in time.

With this in view, the innovative contribution of this paper is that of proposing a unique approach to the design and implementation of all three of the aforementioned services for their use within a museum. In essence, rather than resorting to complex and heterogeneous technologies, which separately address visitor detection/tracking/action recognition, we here propose a unique framework through which all of the abovementioned services can be handled. Not only, we also show that the proposed approach best fits the museum environment building upon the following observation: any employed technology should be non-invasive while being based on off-the-shelf (non-specialized) hardware.

From a technical viewpoint, all this has been achieved resorting to a set of algorithms that can track and recognize the relevant body parts of one or more people, based upon the idea that these can be simply found as the local maxima of segmented foreground areas. Hence, the technical innovation of our work resides in finding for the first time, to the best our knowledge, a unifying and simple approach that can well accommodate the problems that arise when managing human-computer interactions for augmenting the visitors' experience within a museums.

To support such arguments, we describe the design and implementation of two systems that, based on the proposed methodology, have been successfully displayed at a museum (*Mercator Atlas Robot* at the Poggi Palace in Bologna) for one month and at a public exhibition (*Tortellino X-Perience* at the World Expo in Shanghai) for a few days. The display of such performances let us conduct surveys where visitors almost unanimously corroborated with their opinions our initial conjecture: museum interactivity is welcome as long as it is simple and intuitive. This conclusion encouraged us to find new ways of interaction and exploit new methodologies that can extend the number and the quality of the exchanges that can be performed between visitors and exhibits. In particular, when asked to submit a proposal for a new exhibition that will be held at the Puccini Museum of Lucca, Italy, we studied such type of extensions and designed and implemented a system where the actions performed by one or more visitors with multiple parts of their bodies can be followed, as well as recognized.

Summarizing, hence, the areas where this work provides a contribution:

(a) The identification of those services that, based on advanced technologies, are most desired in the context of museums and public exhibitions;

(b) The proposal of a unifying and flexible methodology that can be put to good use to devise algorithms suited to the context of use; all the identified services can be easily implemented utilizing such methodology and off-the-shelf hardware;

(c) The results collected from the deployment of real world systems;

(d) The design of extensions that can also support the recognition of actions performed by multiple visitors with different parts of their bodies (i.e., head, hands, feet).

The remainder of this paper is organized as follows. In Section 2 we present the work that is most closely related in scope to our contributions. We then move on and introduce which are the main means of interactions with a museum in Section 3. In Section 4 we provide more technical details regarding our approach and in Section 5 we discuss the design and implementation of the systems that we deployed at museums and public exhibitions, as well as surveys that witness the validity of our approach. We finally conclude with Section 6.

2. Related Work

The research efforts that have been devoted to all of the possible applications of computer vision are many and articulated. Hence, given the particular context of use, rather than here concentrating on a generic analysis regarding the advantages and disadvantages of different vision-based approaches and solutions (e.g., [7]), we decided to here focus on those technologies that have been so far employed within the context of a museum. In fact, a wealth of research has been carried out and devoted, in the past two decades, to the application of new and advanced technologies in the context of museums and public exhibitions. A large amount of such work did not limit to the sole provision of multimedia content as it is (e.g., display accurate 3D digital reconstructions of past civilizations), but also included the implementation of advanced interaction schemes, laying in front of visitors the chance of participating into digital performances as well as the opportunity of influencing or even deciding what was going to be presented. In the following, we list and briefly describe a few of the systems that have been experimented and

that implement the services that most often are required within the context of a museum: visitor detection, tracking and action recognition.

A wide set of applications can potentially benefit from a service that detects whether a visitor is located in a given area or occupying a certain space. Such applications range from digital tour guides to digital captions, but also include digital menus and public display systems [8], for example, and typically aim at providing engaging experiences and explanations regarding the objects that are displayed close to the area where a visitor is standing [9, 10]. Equally wide is the set of technologies that have been employed, to this date, to provide such type of service [11]. The authors of [12], for example, detected the position of a visitor in a quite unconventional way, utilizing image pattern recognition techniques. In brief, the visitors that entered a museum that employed such type of technology were equipped with a computer tablet and were recommended to take a picture of an artifact when desiring to obtain more pertaining information. A software application running on the tablet exploited pattern recognition and matching algorithms to recognize the artifact, among all those displayed at the given museum, from the image that was captured, thus also locating the position and the orientation of the visitor. The PEACH project [13] exploited a similar approach, as long as completely orthogonal ones, where the authors experimented with infrared devices and radio frequency identification (RFID) ones to detect the presence of visitors at given locations. With infrared technologies, for example, beacons were exploited to establish the distance of a mobile device from a given location. Two known drawbacks accompany the use of infrared signals: their use requires direct line of sight between a sender and a receiver as well as the presence of no interfering obstacles. With RFIDs, instead, transponders and transceivers engage in communications that occur within a given distance, typically a few meters at most, without requiring line of sight, although human bodies heavily attenuate electromagnetic waves and lead to performance degradation in very crowded rooms and different exhibits. Moreover, it has been observed that when two exhibits are placed side by side and are identified utilizing different transponders, which are also associated with different IDs, localizing the exact position of a visitor can be challenging [9].

The problem of tracking a visitor at all times extends that of detecting it at a given location. Differently from the latter problem, the aim here is that of being capable of following a visitor everywhere it goes, outdoor and indoor, within the area of a museum. Outdoor localization is generally considered a non-problem in literature, as global

positioning system (GPS) is widely accepted as an efficient and viable solution that provides position estimates affected by low errors. However, the utilization of such type of technique would generally require a museum to provide its visitors with high-end portable devices equipped with GPSs, while still leaving indoor tracking an unsolved problem. For indoor situations (e.g., inside the Vatican museum Sistine Chapel) WiFi received signal strength indications (RSSI) have been, for example, widely exploited, adopting two basic approaches: one that utilizes the relationship between RSSI and distance and another that builds an RSSI map of indoor space where signal strength values are matched to the locations where they are measured. The authors of [13] employ WiFi technologies in a coarser manner than the ones described, simply utilizing it to detect whether a visitor is located in a given room or not. They implement finer grained localization within a room using infrared or RFID devices, as described in the previous paragraph. Also such type of approach, however, requires the use of some type of device that is WiFi-enabled. A further alternative that has been exploited in literature is based on the use of accelerometers [13, 14]. In practice, building upon the kinematic equation that relates position, speed and acceleration, the authors of various systems obtain position values from acceleration values. Many works, however, also describe how obtaining precise information is rather unfeasible for the precision required in indoor scenarios, as the accelerometer sensors with which mainstream portable devices (e.g., smartphones) are equipped are not accurate and cannot be reliably used for computing accurate position estimates [15].

Fine-grained interactions with the exhibits that are located within a museum, including the recognition of actions performed by visitors on their virtual copies, have been explored in past few years in a variety of different ways [16]. A suggestive stream of work is based on the use of haptic interfaces, where any visitor can perform actions, e.g., touch virtual sculptures, experiencing the perception of ancient artifacts through tactile stimuli provided by force-feedback devices [17, 18]. The author of [19], in particular, exploited such systems to provide physical contacts between visitors and real sculptures, otherwise prohibited because of their preservation. A completely different approach, instead, is instead exploited in all those systems that are based on custom accelerometer sensors [20], where accelerometers and hidden Markov models (HMM) are put to good use to recognize a set of basic input gestures that a visitor performs with its hands.

Now, within all the previous experiences that we here listed, a general trend can be found: all of the solutions that we have discussed so far require visitors to use some type of hardware device. This aspect is key for many different public exhibition environments, as other means of mediation between visitors and exhibits can also be exploited and used. Upon such grounds, we decided to embrace a philosophy where a visitor is portrayed as a free actor moving throughout a museum, without the need of any electronic device. All the sensors that are required to recognize and interact with a visitor are placed within the museum. A number of systems have already put in practice such idea, avoiding the provision of any hardware equipment to visitors [21-29]. One of the pioneering works in such field dates back to the '80s, when computer vision techniques were first exploited to recognize the whole body of a player as an input device to a videogame [30]. More recently, the authors of [31], for example, adopt similar ideas utilizing cameras for the recording of the movements and actions. More specifically, in their work they define a distinctive form of augmented reality, i.e., social immersive media, where multi-user interactive camera/projector are put to good use to promote social interactions among visitors and exhibits. For example, they employed such type of paradigm when, in the *Fear* performance, they exploit video cameras to recognize the movements performed by a visitor, while immersed in a virtual world, the Amazon forest, where any incautious movement can be recognized by a hungry (virtual) jaguar. An equally interesting support of interactivity can also be found in Simpson's Shadow Garden [32], where the shadows produced on a projection screen by visitors are captured and exploited to interact with virtual objects and characters as water and butterflies.

Although all of such approaches point in the direction of providing natural and intuitive means of interactions between visitors and digital performances, we believe that more can be done when considering both of the actors that play a primary role in any museum ecosystem: the visitors and the museum itself. With such awareness it is possible to narrow down those services that are effectively needed at an exhibition site. In fact, we observe that the solutions that are often encountered, are often borrowed from other applications (e.g., gaming industry), and not tailored to the specific conditions that are instead found within the context of interest.

3. Body-based Interactions within a Museum

Museum guides act as mediators when they play one of the following roles: (a) they select the exhibits they deem

worthy of attention, (b) they disseminate precise information about the exhibits, and (c) they interpret the exhibits, translating any unfamiliar messages they convey to their visitors [33]. While these tasks were once exclusively carried on by human guides, they can now also be accomplished adopting interactive technological solutions. Focusing now on how all of the aforementioned mediation roles can be efficiently supplied in such type of environments, we here aim at jointly finding: (a) the interactions that are more often desired in such type of spaces, as well as, (b) the technologies that can implement those interactions, in the most discreet and invisible way.

As anticipated, the interactions that are most often sought for concern three types of services: visitor detection, visitor tracking and visitor action recognition. Very often, as analyzed in Section 2, such services have been offered utilizing tangible and visible hardware devices. Now, we here lay our eyes on the characteristics of the two actors that play within such context (i.e., the museum and its visitors) and derive the technological requirements of the systems that can best match their needs.

In fact, from the perspective of a museum, any employed hardware devices should be:

(a) Noninvasive, as often the buildings and environments where museums are hosted are part of the exhibition itself (e.g., Palazzo Strozzi in Florence) and hence not changeable at will;

(b) Flexible and easy to move, as often exhibitions are temporary and are characterized by a time horizon beyond which an artifact may be transferred or even become inaccessible.

Interestingly, a thorough analysis of the state of the art reveals that also visitors share similar requirements: it is widely accepted that the interfaces that support the most natural (e.g., gestures in the free air), but even familiar (e.g., smartphones), human-computer interactions are those that are built upon those technologies that either reduce the use of any hardware device or are based on technologies that are well known to the general public [7, 24, 31, 34, 35].

Concluding, the aforementioned arguments should convince the reader that there is room for the use of body-mediated interactions within the context of museum. The next Section will be devoted to discussing how body-mediated interactions can be put to good use in this setting.

4. Intercepting Interactions within a Museum

While our aim here is not that of refuting the use of any legacy technology (and the forms of mediation that they support) that has been put to good use in the context of study, we will here provide a detailed discussion that is centered around the adoption of a unique approach to support the interactions and services that have been so far individuated.

4.1. General Considerations

In this Subsection we will present the conditions that should be met to support body-mediated interactions in the context of a museum. The best results, along the lines of supporting body-based interactions, can be achieved when museums are equipped with the least number of sensors that can *pull* information from visitors, who never hold or carry any hardware, but simply interact with the artifacts that are displayed moving and performing gestures, freely, in the air. Such types of interactions [36], where no physical contacts occur (i.e., intangible interactions) are very convenient as long as they prove to be: (a) intuitive, and, (b) accurately identified.

All this provides us with a hint regarding the direction that should be taken when instantiating the desired visitor detection, tracking and action recognition services. We here pick up this tip and identify three different instances, which can all be realized with the sole use of one or two webcams feeding real-time video streams to efficient software algorithms: (a) intangible detection, (b) intangible tracking, and, (c) intangible action recognition.

With intangible detection we circumscribe all those interactions that are triggered when a visitor (i.e., its blob) occupies a given space (i.e., location on a frame). This is the simplest case, as it only requires the definition of a set of checkpoints, e.g., *sensible areas* that, when occupied, generate some type of response. In such situation, hence, the frames that capture a visitor's body do not undergo any complicated processing operations, as no segmentation procedures and no representative body points are computed.

We talk of intangible localization when every single visitor is associated with a point at all times, i.e., its location. Contrarily to the intangible detection situation, bodies are here segmented and a representative point is computed to identify each detected blob, which is continuously followed, as long as it moves within the area where the service is

active. In practice, such service achieves the same result that could be attained with a GPS, with the important difference that positions are determined without requiring any external device.

Finally, we talk of intangible action recognition in all those situations where pre-defined actions performed by visitors can be recognized. Such type of service is the building block for all those performances that trigger some type of reaction as a consequence of a given action. Clearly, the design of such type of interactions requires the recognition and tracking of more than one point, as an action can origin from any part of a human body and can be as articulated as a sophisticated movement of a hand (e.g., sewing pants), an arm (e.g., maneuvering an oar), or also a foot (e.g., kicking a ball). In such case, then, bodies cannot just be segmented or represented by a single point, but some type of efficient algorithm is needed to track all relevant points (e.g., head, feet and hands) at all times.

In Figure 1 we summarize the main characteristics of the three services, emphasizing the geometrical entities that they follow (areas, a single point for each visitor, as long as multiple points for each visitor).

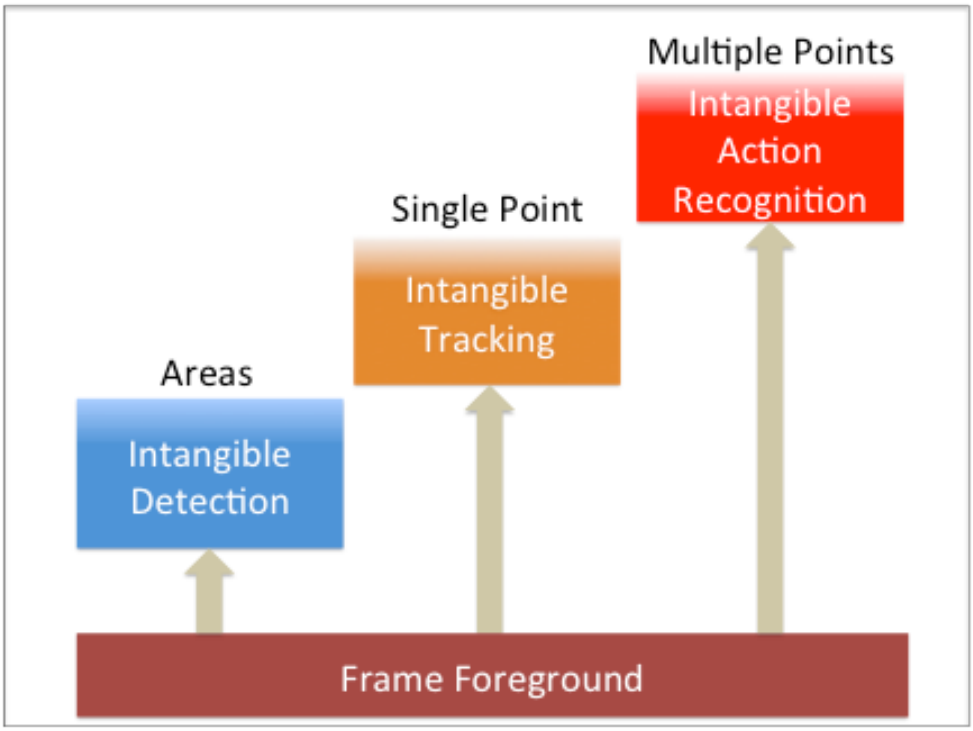

Figure 1. Intangible services: increasing complexity and number of geometric entities that are followed.

In Table 1, instead, we outline the main characteristics of the three services that we have so far described distinguishing them as a function of: (a) the number of relevant points that they can detect, as long as, (b) the type of interactions that they support. In addition, for each service we also provide the reference to an exemplar performance system that adopted it.

Table 1. Museum services, tracked points, function and reference example.

Service Name	**Number of Points**	**Type of Interaction**	**Example**
Intangible Detection	No points	Touched Area/Space	Snibbe and Raffle 2009, Marfia et al. 2012
Intangible Tracking	Single Point	Follow a single point within a museum	Malerczyk 2004, Shridar and Sowmya 2008
Intangible Action Recognition	Multiple Points	Follow multiple points within the performance area	Roccetti et al. 2010

4.2. Intangible Interactions: A Local Maxima Approach

Now, we have arrived to the point where we want to know how these concepts can be put in practice, thus searching for viable answers to the following question: how can we devise simple and efficient systems that implement all of the aforementioned operations with one or two webcams at most? In the following we answer such question providing a novel methodology, that leads to the design of algorithms that well suit the different contexts that can be encountered within a museum: detect whether an area has been traversed, follow a single point for each visitor, providing multiple points for each visitor, respectively. Hence, we proceed describing the steps on which such methodology is based. The first step entails understanding how the blob that identifies a human body can be circumscribed.

This first operation is at the basis of both the intangible detection service, which worries about revealing whether some area has been *touched* by some blob, and the intangible tracking and action recognition services, which extract relevant points from the blobs that are provided. Algorithms that implement such type of operations have been for long studied and are readily available [38, 39]. The simplest ones are those that adopt image background subtraction techniques: foreground objects are detected as the difference between the current frame and a frame that represents the static background. Hence, the problem of detecting whether a certain area has been touched (i.e., intangible detection) simply reduces to that of checking whether any foreground object can be identified at that given location within a frame.

The intangible tracking and action recognition services, instead, require second step which is formed through filtering processes that lead to the derivation of relevant points (i.e., a single point for each visitor when tracking, multiple ones when recognizing an action) from detected foreground areas. Such filtering processes represent a transformation $\mathscr{T}$ that, based on geometric considerations, takes as inputs all foreground pixels and delivers an estimate of all sought points, as a function of the following variables: (a) the expected posture of a visitor (e.g., sitting, standing, etc.), (b) the relevant points themselves (e.g., seeking for only the head of a visitor, or also for his/her hands), (c) the characteristics of the areas that are under analysis (e.g., illumination level), (d) the positions of the webcams (e.g., placed above or on the side), and (e) the adopted filtering principle. While variables (a) to (d) are all clear, as they all pertain to the physical properties of the area as well as to the properties of the relevant points that are sought, a few more words are needed to explain what variable (e) is. The filtering principle, in fact, is the algorithm that is adopted to detect relevant points from the pixels that compose the foreground areas. While many different principles have been proposed in literature in the past years, we here keep consistent to an approach that seeks for solutions that are simple and flexible. Hence, we here propose a methodology that can be applied to many different situations, which is that of detecting all relevant points exploiting the cascade of two different schemes (bottom block diagram of Figure 2).

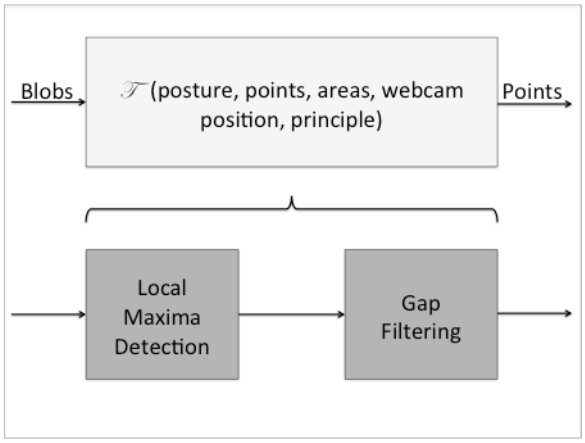

Figure 2. Finding relevant points in blobs.

The first scheme (Local Maxima Detection in Figure 2) is in charge of detecting all the Local Maxima (LM) that can be obtained from the detected blobs. Such LM are potential points of interest, i.e., a point that is detected during this phase could be a relevant point, but could also represent a false positive (a point that is found in this phase that is not amenable to any sought body part).

The second scheme (Gap Filtering in Figure 2) is used to prune any detected false positive. Anatomical considerations let us assume that any two given adjacent maxima representing relevant points will always be separated by at least one minimum value (i.e., a gap) that is below a given threshold, whose value depends upon the points that are sought.

Now, let us sketch a model of how things work (anticipating here the principles that lie underneath the algorithm that is presented in Subsection 5.3), for example when a camera is placed in front of a scene where the head and the hands of a person could be detected. In such case, the problem reduces to searching for the coordinate values of the pixels (i_L, j_L), (i_R, j_R), (i_H, j_H), which locate, respectively, the left and right hands and the head when a blob, $B=\{(i, j)\}$ (i.e., represented here by a succession of pixels (i, j)), is detected in front of the webcam. We can find such points as follows. The first step amounts to detect the contour line that upper limits the body: this can be done finding the set of local maximum (*LM*) points, where *LM* includes all the local maximum of blob *B* in correspondence of a given abscissa value *i*, as described in Equation 1:

$$LM = \{(i, j) \mid j = max_B(i), \text{ for each } i\}. \quad (1)$$

Once the set of local maxima *LM* is known, the following procedure can be adopted to individuate which pixels in *LM* effectively count. In fact, it is possible to find all the relevant points of interest combining two observations. The first is that an element of interest in *LM* certainly exists and can be readily found: this is the global maximum, (i_M, j_M), of the contour (i.e., the global maximum, for a person that is standing in front of the webcam, can not but be representative of the head or one of the two hands). Now, in case another relevant point is present (i.e., there might be only one), we should factually consider that it would be separated from another maximum by a gap (i.e., a local minimum). Hence, starting from the abscissa where the global maximum is located, i.e., i_M, and proceeding in the positive and negative directions, we can iteratively search for any other relevant point as follows. Let *f* be the function that returns the ordinate value of a point on the contour in correspondence to a given abscissa (i.e., $LM = \{(i, f(i)) \mid \forall i\}$), so that the global maximum may be obtained as $j_M = f(i_M)$. Now, assuming we iterate in the positive direction (the same would work moving in the negative one) from where the global maximum is located, we find that a gap may exist if we are able to find an abscissa i_m where the following holds:

$$i_m = \min\{i\} > i_M, \qquad \text{s.t. } f(i_m) < y_M - GAP, \quad (2)$$

where *GAP* here represents a threshold value that reasonably accounts for the distance that vertically separates the head of a person from his/her shoulder. After that the position of a gap has been found iteratively applying the conditions summarized in Equation 2, we can now seek, moving in the same direction, for a local maximum. A first candidate point can be found as shown in Equation 3:

$$i_{LM} = \min\{i\} > i_m, \qquad \text{s.t. } f(i_{LM}) > f(i_m) + GAP, \quad (3)$$

and can be updated as long as a point $i > i_{LM}$ exists where $f(i) > f(i_{LM})$ holds, before encountering another gap. The procedure valid for finding all the relevant gaps in the positive direction is summarized in Table 2. In particular, in line 1 the algorithm sets the starting point to where the global maximum was found. Proceeding then in the positive direction (the same specular procedure will be also applied in the negative one), the algorithm cycles between the position where the global maximum was found and the maximum abscissa value in the frame (MAX). Condition at line 3 will never be satisfied, until a gap is first found and stored in y_m (line 5). When this condition is met, the position of another gap is stored (*store*(*i*, *f*(*i*)) at line 6). The algorithm proceeds searching for all the gaps and

storing them. Once it terminates, it is easy to find all the local maxima that lie between the individuated gap regions. In order to further explain how the algorithm reported in Figure 3 works, let us now consider, for simplicity, a situation where the posture of a visitor is known in advance and corresponds to the one shown in the top images of Figure 4. Assuming also that a single webcam is placed in front of that given visitor, the problem involves finding the positions of the right hand and the head of the visitor. The position of the hand, in the given situation, can be immediately detected from the foreground blob (top-left part of Figure 4), as anticipated before, as the position of the maximum point in the upward direction (top-right part of Figure 4). The problem, hence, now reduces to finding the position of the head. This cannot clearly be selected choosing, for example, the second maximum value, as that may represent the thumb of the same hand that was previously detected.

1. $y_M = j_M, y_m = j_M, i = i_M;$
2. $while$ $(i <$ MAX) {
3. $\quad if$ $(f(i) > y_m + GAP)$
4. $\quad\quad y_M = f(i);$
5. $\quad if$ $(f(i) < y_M - GAP$ OR $f(i) < y_m)$
6. $\quad\quad y_m = f(i);$ $store(i, f(i));$
7. $\quad else \quad y_m = y_M;$
8. $\quad i = i + 1;$}

Figure 3. Gap filtering algorithm for the positive direction.

As anticipated, such problem can be easily solved, instead, observing that a gap (i.e., a minimum value) always exists between the two maxima that represent the hand and the head in such situation, unless the hand joins the head, situation yielding a single maximum with this methodology. In conclusion, it is possible to distinguish the points that are relevant from those that are not (bottom part of Figure 4) simply implementing a filtering procedure that accounts for the gap that is typically observed between the two maxima that represent a raised hand and the head of

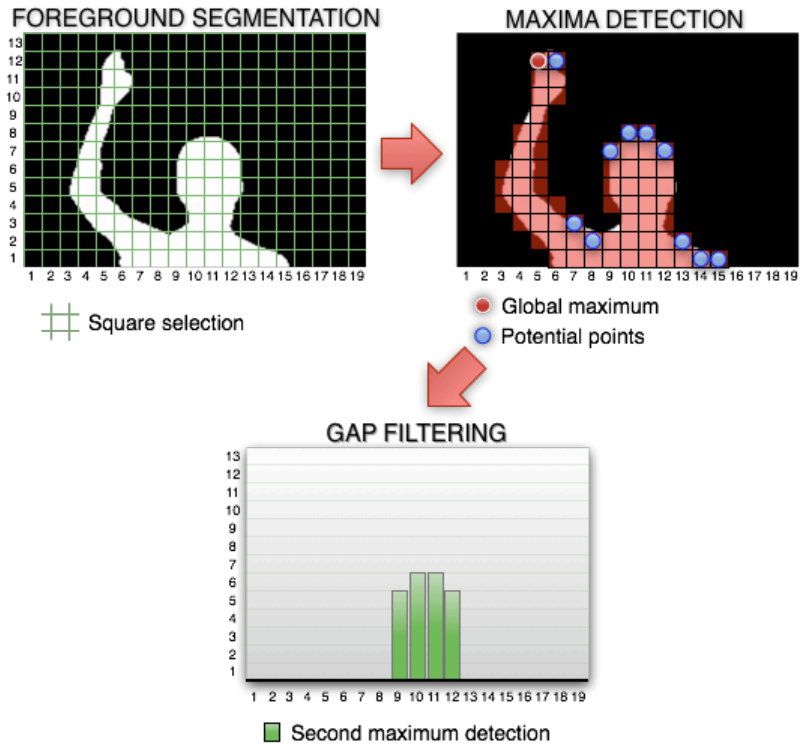

Figure 4. Detecting the positions of the hand and of the head.

a given person, respectively.

Now, assuming we here set the *GAP* value equal to 3 and running the algorithm shown in Figure 3 on the segmented area presented in Figure 4, we find that at the first iteration $y_M = j_M = 12, y_m = j_M = 12$ and $i = i_M = 5$. We unroll all the steps that take to the individuation of the gap areas adjacent to the head in Figure 4 in Table 2. The very final step then amounts to find the maximum value between all individuated gaps (bottom part of Figure 4).

More than mere theoretical speculations, all of such ideas have been stressed and put to good when devising two digital performances that have been displayed in public exhibitions: *Mercator Atlas Robot* and *Tortellino X-Perience*. In the next Section we move on to proceed showing how this approach, i.e., the use of intangible services in digital performances, has benefited their implementation. In particular, we will focus on showing: (a) how these applications have been accepted by the public, and, (b) how our approach can be extended to more complex scenarios.

Table 2. Algorithm simulation run.

i	***f(i)***	y_M	y_m	***operation***
5	12	12	12	-
6	12	12	12	-
7	3	12	3	*store*(7, 3)
8	2	12	2	*store*(8, 2)
9	7	7	7	-
10	8	7	7	-
11	8	7	7	-
12	7	7	7	-
13	2	7	2	*store*(13, 2)

5. Intangible Interactions: A Few Case Studies

In the following we present and analyze the results drawn from three case studies, where intangible interactions were supported within the context of a museum or a public exhibition. All case studies reflect the primary scope of this work, which is that of proposing new means of mediation between a visitor and an artifact, based on a visitor's body and on the effects that can be triggered with its movements. To this aim, we here focused on how the relevant features of the human bodies were detected in those regions of space that enclosed both the visitors and the exhibits, rather than on how the positions of visitors could be followed at all times [24]. In this way we define a sort of new and dynamic environment that appears every time a visitor and a displayed exhibit get together in these regions. Under these conditions, we here demonstrate the validity of the proposed service model and highlight the importance of its intangible principles, as well as the importance of the methodologies and the algorithms that we devised to support them. In particular, the first case study, *Mercator Atlas Robot*, is a digital performance that accompanied the exhibition of an ancient book (Mercator's printed Atlas) for more than one month at the Poggi

Palace Museum of Bologna, providing a representative exemplar of those performances that are based on some type of intangible detection of visitors [37]. The second case study describes *Tortellino X-Perience*, a video game that has been displayed at the 2010 Shanghai World Expo and that implemented an intangible recognition of the actions performed by visitors [34, 40]. We finally conclude with a case study that demonstrates how the recognition of more complex actions can be supported with simple extensions of the same design principles that have been presented so far.

5.1. Mercator Atlas Robot (2012)

Mercator Atlas Robot is a digital performance that has been active for over one month at the Poggi Palace Museum of Bologna, Italy. Its purpose has been that of serving the exploration process of an ancient and important volume that has recently been rediscovered in one of the University's libraries. The volume, printed in 1630, is one of the last remaining copies of the 10th edition of Mercator's World Atlas [37]. Interested at how intangible mediations could be adapted to fit such scenario, as well as accepted by the general public, we will here focus our discussion around its design and the assessment of its performance through a set of survey results, with a larger scope and a more in depth discussion that that provided in [37].

While the importance of this finding was immediately recognized, museum curators also readily understood that a volume of such beauty (composed of over one-hundred gorgeous maps) and scientific relevance could not be simply buried in a crystal case, leaving no consultation opportunities to museum visitors. This situation, hence, offered us the opportunity of being invited to be part of a team, which also included a historian, an art director and a museum curator, that worked at finding a way of letting visitors peek into the Atlas, truly enjoying its content, without putting at risk its preservation. The solution, in simple words, was found with the design of a digital performance system, termed *Mercator Atlas Robot*, which projected the pages of the volume on one of the walls of the museum. Unlike a mere projection, however, *Mercator Atlas Robot* is, in fact, a truly mixed reality system that supports a transparent navigation of the ancient maps that are contained in the Atlas in a novel and immersive way. Such result has been attained as follows.

In fact, reinventing in contemporary terms the metaphor initially thought by Mercator (an engraved Atlas holding and inspecting the world on the frontispiece of his work), we repeated it placing our visitors, instead of the Titan, in the position of holding the world while raising their arms and exploring the book's maps. This was implemented positioning a hidden webcam in front of visitors and placing a set of intangible buttons (e.g., Virginia and South America in Figure 5) above the open virtual book that displayed Mercator's maps (Figure 5). A visitor, while observing the displayed map, could browse the book simply *touching* one of such buttons, e.g., Virginia in Figure 5,

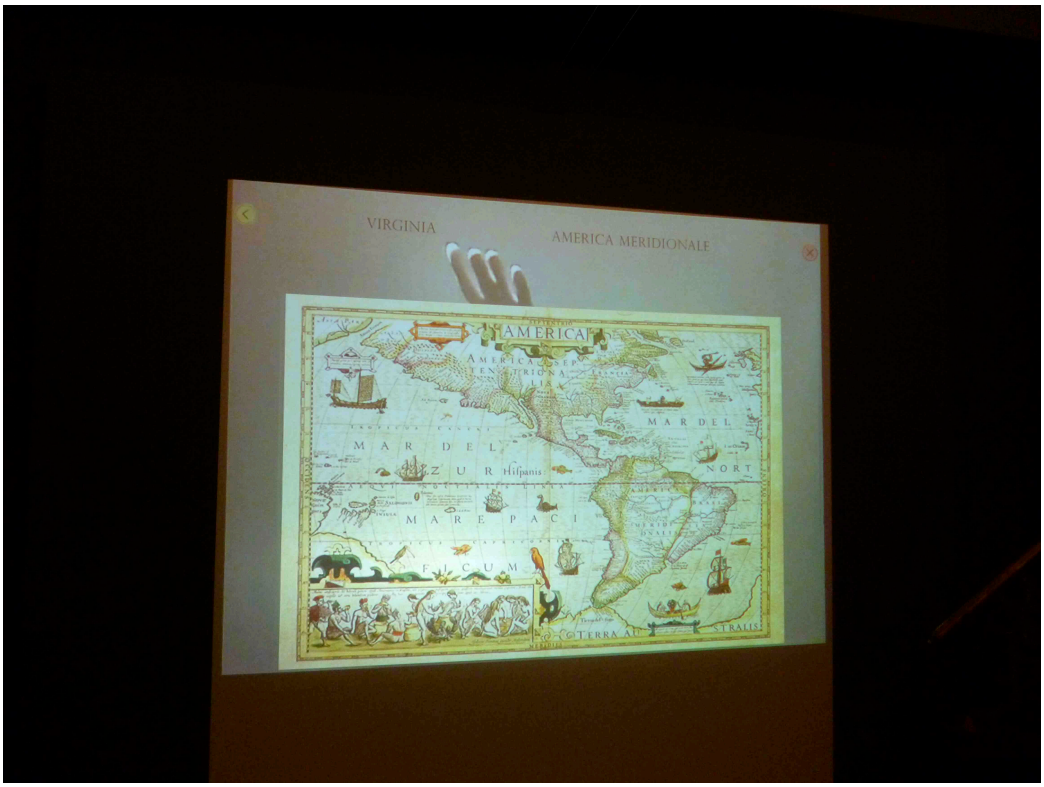

Figure 5. *Mercator Atlas Robot* mixed reality performance.

ending to be immersed in the corresponding map. Clearly, all this was possible thanks to the design of an intangible detection service that sensed when a given virtual button was activated.

Now, rather than providing further details on how the graphical environment of *Mercator Atlas Robot* has been implemented (further information can be found in [37]), we are here more interested in providing the results from the survey that has been administered to the visitors that employed this digital performance. In particular, we asked

four detailed questions, that are listed in Table 3, and also posed an open one, that visitors could answer freely, that sounds as follows: “Tell us what you enjoyed most”. The first question, “Have you ever experienced such type of performance?”, could be answered either YES or NO. Out of two hundred and forty four completed surveys, we found that only a few visitors responded positively, yielding that almost 80% of visitors experienced their first performance of such type when interacting with *Mercator Atlas Robot*. The second, the third and the fourth question could be answered, instead with a number, ranging between 1 and 5, where 1 meant “low” and 5 “high”, inspired by the *Likert* scale [41]. In Table 3 we report the average, along with the minimum, maximum and standard deviation, where applicable, computed over the whole set of values indicated by visitors. In particular, average answer values that exceed 4 for both the second and the third questions point that there has been, in general, a high appreciation of the performance and that the visitors regarded it as highly intuitive and easy to use.

Table 3. *Mercator Atlas Robot* survey analytic results.

Question	**Answer**	**Std. Dev.**	**Min/Max**
Have you ever experienced such type of performance?	NO: 78%	----	----
Did you enjoy it?	Average score: 4.11	1.15	1/5
Was it easy and intuitive to play?	Average score: 4.24	1.03	1/5
Would you have preferred playing with a remote control?	Average score: 1.89	1.29	1/5

Oppositely, an average answer value below 2 for the fourth question suggests that the means of interaction that have been adopted, i.e., intangible interactions, have been highly regarded by users (visitors would have NOT preferred the use of a remote control).

Of particular interest are also the answers that visitors provided to the last open question. In fact, most of such

answers fit in one of the two following groups (Table 4). The first group (over 20% of answers), which is the one that also contains the greatest number of answers, includes all the feedback indicating the appreciation for the performance as a bridge towards a world that could not be otherwise visible: the ancient volume. This is clearly very important, as it witnesses that *Mercator Atlas Robot* has not been experienced as an actor of the museum itself, but as an invisible link (i.e., mediator) between the exhibit and its visitors (hence avoiding any Guggenheim effect risks). In the second group (almost 10% answers), instead, we included all those visitors that posed their attention and appreciation to the chosen means of interaction with the exhibit. Although this was not the final aim of our work, this feedback is also very relevant, as it corroborates the idea that intangible ways of interaction provide more natural and more comfortable tools to play with any object, be it a smartphone or an ancient book.

Finally, we strongly feared during the design stage of this performance that the novelty of the proposed interaction scheme would have prevented visitors from participating with our performance. Interestingly, as it is witnessed by

Table 4. *Mercator Atlas Robot* open question.

Open Question	**Percentage**
Appreciation for the performance as a bridge towards a world that could not be otherwise visible	20%
Appreciation for the chosen means of interaction with the exhibit	10%
Generic considerations without any direct relationship with the installation	70%

the survey results that we here reported, as well as informally by the staff that constantly followed the performance in the museum, this was not the case. What, instead, positively surprised us was that *Mercator Atlas Robot*: (a) attracted both children and elders, and (b) facilitated not only gestural interactions, but also social ones, because of the curiosity that it raised.

It should be clear that while Table 3 provides a concise representation of our analytic survey comprised of specific

questions that could be answered utilizing a 1 to 5 scale of values, Table 4, instead, reports on the part of the questionnaire which was left open to general comments of visitors. The rationale behind the first part of the survey was to steer the users to assess specific features of the system, while the second part was, on the other side, completely open. In conclusion, while we concur on the fact that measuring the user experience of a system could entail engaging more sophisticated technologies, we cannot neglect that we dealt with a very stringent environmental factor: our system was displayed during a real exhibit. This means that our questionnaire was not supplied online to professionals, but, on the field, to the general public (which includes kids and elder people), that was generally not happy of answering to complex surveys. This motivates why we searched for a trade off between the scientific completeness of the survey and the likelihood of real people answering to its questions.

5.2. Tortellino X-Perience (2010)

Tortellino X-Perience is a digital performance that has been displayed for a few days at the Shanghai World Expo in October 2010, at the Cantina Bentivoglio in June 2011 and at the Bologna city hall in September 2011 [42-45]. Although it has never been exhibited, so far, within the walls of a museum, all the mentioned environments where it has been shown very closely resembled the typical situations, and problems, that could be found in cultural exhibitions. The Expo, for example, closely resembled a museum, offering a scenario where again an exhibit, a visitor and a mediator got together and interacted. Just as it is typically done in museums, *Tortellino X-Perience* has been played in a kiosk set in an open space, where a new player could join every few minutes. The scope of such performance was that of teaching how the original Tortellino pasta (i.e., a traditional Bolognese pasta dumpling) is prepared. This was accomplished challenging players at copying the actions that real cooks perform when preparing Tortellini.

Playing essentially amounted to going through a number of phases, where each phase was organized as follows: (a) learn cooking gestures while watching a video, (b) reproduce those actions, and, (c) watch the result. All this was accomplished displaying an immersive environment in front of players (leftmost part of Figure 6) and taking advantage of an intangible action recognition system. In fact, a video camera was placed above players (rightmost

Figure 6. *Tortellino X-Perience* in Shanghai.

part of Figure 6), serving the purpose of capturing their hands, whose positions were then located and followed by a set of software algorithms. Hence, if the actions that were reproduced by a player were incorrect, that player was then asked to repeat the gestures correctly, until the right sequence of movements was performed.

Now, *Tortellino X-Perience* represents a relevant exemplar of a system that employs an intangible action recognition service. In particular, the service that this performance utilizes has been devised to solely detect the points that represent the two hands. Hence, when searching for the two hands of a visitor that is extending its arms within a frame that has been captured from the overhead position, a first hand can be immediately located at the maximum foreground point in a given direction (bottom-left part of Figure 7).

The problem, then, reduces to finding the position of the second hand. This cannot be selected choosing, for example, the second maximum value, as that may represent the thumb of the same hand (same problem encountered in Section 3). Such problem can be easily solved observing that a gap (i.e., a minimum value) will always exist between the two maxima that represent the two hands, unless the two hands overlap. Implementing, therefore, a filtering procedure that is based on the gap that is typically observed between the two maxima that identify the edges of two extended arms (i.e., the two hands), it is possible to distinguish points that are relevant from those that are not, and hence precisely identify the two hands (bottom-right part of Figure 7).

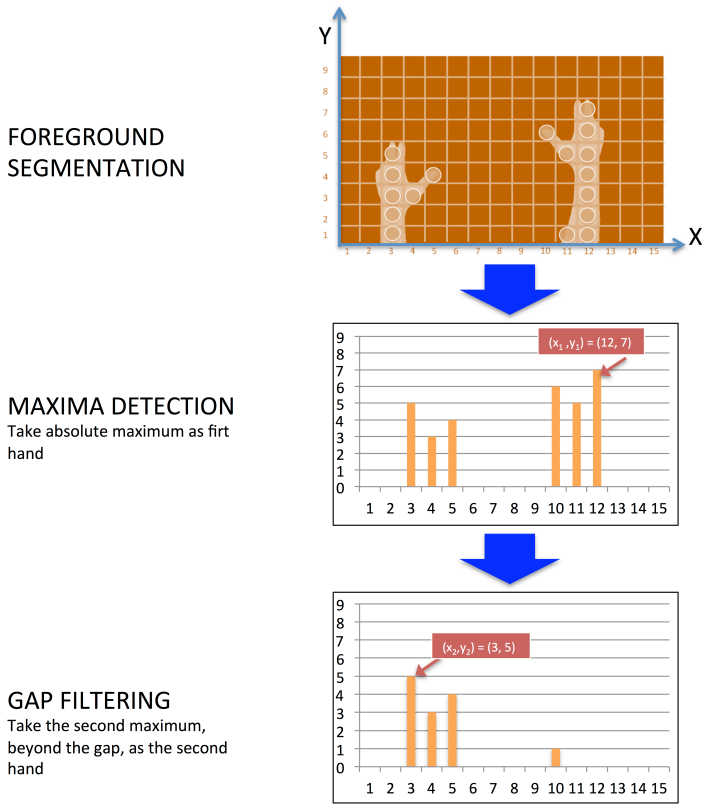

Figure 7. Finding hands in *Tortellino X-Perience*.

Once the two hands are detected, within each frame, the specific actions that they perform can be recognized a sequence of areas touched within certain time constraints.

We are, instead, interested, at this point, at understanding the level of appreciation that a performance like *Tortellino X-Perience* achieved (Table 5).

Table 5. *Tortellino X-Perience* survey results.

Question	**Result**
Did you enjoy the game?	Average score: between 4 and 5 (max = 5)
Was it easy and intuitive to play?	Average score: between 4 and 5 (max = 5)
Would you have preferred playing with a remote control?	NO: > 90%

Also in this case we laid forward a set of questions that a good number of participants (over a hundred) answered [34]. The questions were a subset of those that we later posed for *Mercator Atlas Robot*, and sounded as follows: (a) "did you enjoy the game?", (b) "was it easy and intuitive to play?", and, (c) "would you have preferred playing with a remote control?". Similarly as before, players were asked to rate, assigning a score between 1 and 5, to the enjoyment and the comfort that they experienced while playing, when answering the first two questions. The third question, instead, only permitted a positive or negative answer in this case. Interestingly, the results that we collected for *Tortellino X-Perience* have been later corroborated by *Mercator Atlas Robot*: in general players enjoyed the Tortellino game assigning an average score above 4 to the first question, while they also found it easy and intuitive to interact with, as an average score well above 4 also confirms. Finally, players again preferred being free while interacting with the game, as the great majority of answers where negative to the last question.

Unfortunately, we did not give the chance to visitors to provide open feedback evaluations, which, as *Mercator Atlas Robot* later revealed, provide valuable and insightful information regarding how the general public perceived the performance.

5.3. Extending the Methodology: The Case of the Puccini Museum (2012)

We will here report on a feasibility study conducted for the renovation of the Puccini Museum that is based in Lucca, in the house that gave him birth. In order to be able to support a wide set of physical interactions, we devised a new algorithm that serves the purpose of identifying five relevant points of the human body, at any given time. Following our example, the reader can get convinced that the sole use of these five points is key to implement a mediation process between a visitor and an artifact in that given museum. In particular, in the remainder of this Section we will demonstrate the extent to which it is possible to support complex and imaginative interactions with objects and exhibits, that in this particular case, include both pictorial content (e.g., photos, musical scores, etc.) as well as music instruments (e.g., the instruments utilized by Puccini while composing the *Turandot Opera*). The demonstration is carried out assessing our approach on a benchmark composed of 57 possible body positions. Clearly, those 57 only span a broad gamut of all the possible gestures that can be supported (potentially infinite).

In this Subsection, hence, we will focus on the design and experimentation of a system that, utilizing a single frontal webcam and a single overhead one, tracks the position, feet, hands and head of one or more people. The imaginary situation depicted in Figure 8 (we did not have the honor of working for Leonardo's masterpiece, so far, but we are, indeed, very much honored of working to keep alive the heritage of Giacomo Puccini) is representative of the considered scenario: two visitors move in front of an exhibit (e.g., the *Monnalisa* in the leftmost part of Figure 8) while their positions and their relevant points are detected in real-time (rightmost part of Figure 8).

All this is possible exploiting the frames that are streamed in real-time by two webcams, one that is placed above the visitors (clearly distinguishable in both of the images shown in Figure 8) and a second that sits in front of them (above the painting, in the leftmost part of Figure 8). The frames captured by both cameras are utilized for intangible tracking operations, as the combination of the information that they provide can be put to good use to find a representative point that identifies each player. The approach that we exploit is again that of first finding a maximum value in one direction, identifying hence the first player, and then deriving the second one beyond a gap deeper than a given threshold (this same procedure can be applied to find the positions of any arbitrary number of visitors).

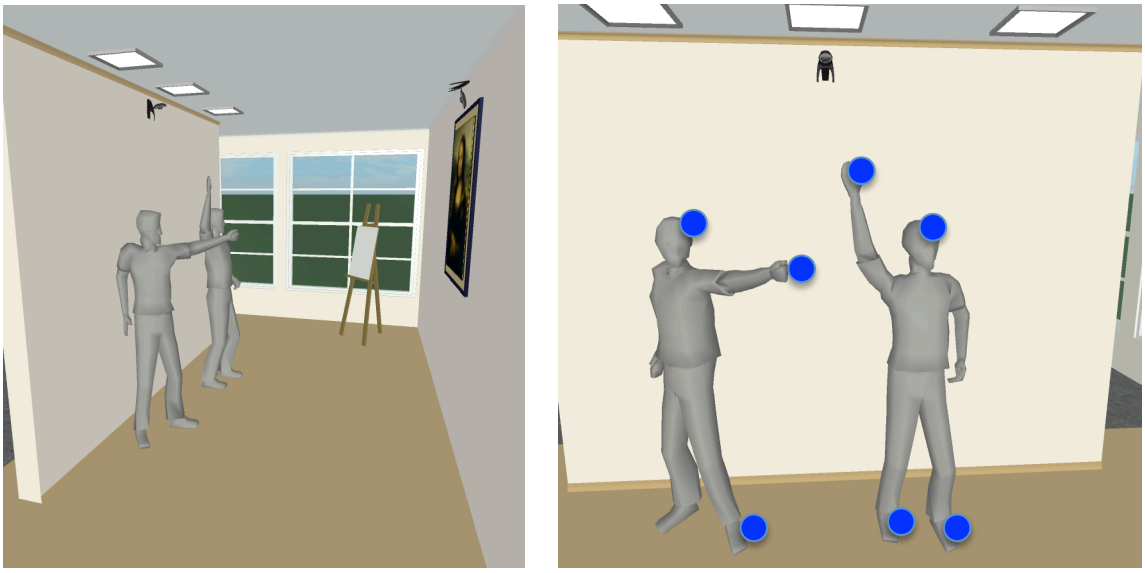

Figure 8. Two visitors interacting with the *Monnalisa*.

Now, what we are most interested at is to show how it can be possible to detect multiple relevant points processing the frames captured from a single frontal webcam, nonetheless giving a detailed description of all players that fall into the webcam's field of view at any given time. Assuming that the positions of players are given by the preceding procedure, we can now concentrate on the blob of each single person in isolation and find where the head, the feet and the hands lie, just as shown in rightmost part of Figure 8.

The methodology that we here propose is capable of understanding whether a body part can be reliably identified, or not. In case it cannot find a reliable point, it will simply return no information (just as it happens for one of the two visitors in Figure 8). It hence classifies detected points as: (a) accurate, and, (b) approximate ones. Accurate points are those that have been found with a very good precision, as they are always very close to the body part that is being followed, while approximate points are the result of guesses based on anatomic observations. As we shall see from our results, almost all points that are found by our algorithm belong to the first category.

Our algorithm is composed of five major steps, that can be distinguished in Figure 9: (a) lists population, (b) proportions extraction, (c) feet detection, (d) top-down peaks detection, and, (e) remaining hand-points detection.

With the term list population we here call the procedure with which we:

- Select all those squares that are detected in the foreground of a frame;
- Store their positions in, respectively, two lists, one where they are sorted from the top to the bottom and another where they are sorted from left to right.

The rationale of such step is that of keeping a data structure that serves the purposes of supporting: (a) direct access (thus requiring constant computational time) to the positions of all the points that compose the contour of a blob, and, (b) fast operations (e.g., distances) among points (e.g., adjacent points on the blob are also adjacent on the list).

The second step, termed proportions extraction, is a divide and conquer phase that serves the purpose of distinguishing the different areas of a human body (i.e., top from bottom, left from right). In particular, utilizing anatomic proportionality rules it allows distinguishing:

- The lower from the upper part of a blob, with a horizontal axis drawn at one-third of the height of the entire blob;
- The upper-left part from the upper-right one, with a vertical axis that splits in half the upper part of the blob;
- The lower-left part from the lower-right one, with a vertical axis that splits in half the lower part of the blob.

The first relevant points that are extracted from a human blob are the two feet. This step is performed finding the global minimum in the lower part of the blob, point that will represent the first foot, and searching for the second foot in the left or right regions, depending on where the global minimum was found. The position of the second foot is hence obtained as the local minimum beyond a maximum point (i.e., gap) that exceeds a given threshold (just as in the case of *Tortellino X-Perience*, when searching for the positions of the two hands).

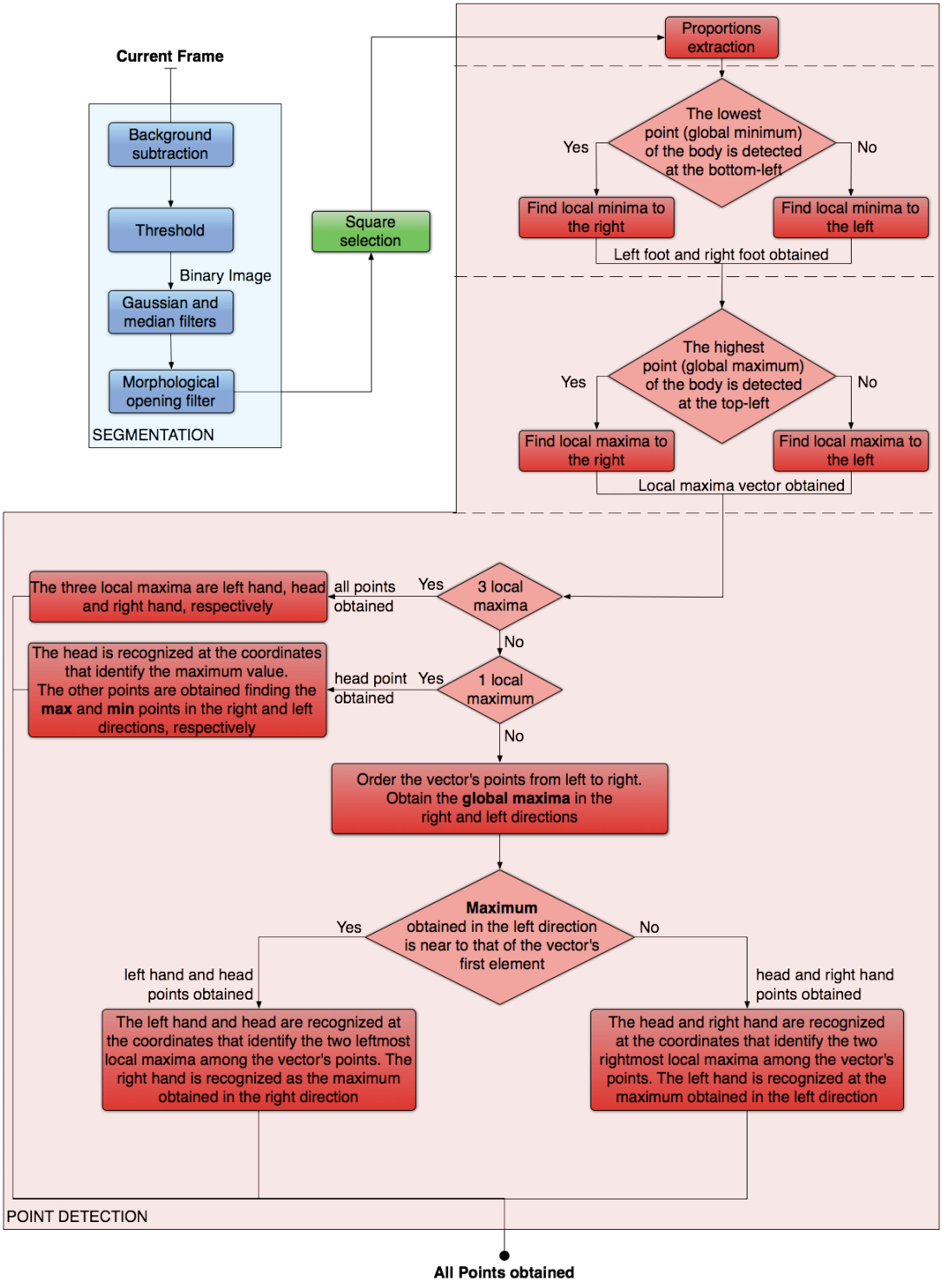

Figure 9. Algorithm flow chart.

The top-down peaks detection step is the one that is responsible for searching, again exploiting maxima and minima values, the relevant points that belong to the upper part of the body (i.e., hands and head). The rationale here is that of first checking whether a single global maximum or three local maxima can be immediately distinguished. The former situation indicates that the head has been found. The remaining other two points (i.e., left and right hand) are obtained in the successive phase, i.e., remaining hand-points detection, as the absolute maxima in the left and right directions, respectively. In the latter case, instead, the algorithm was lucky enough to find right away the positions of both the hands and the head of a visitor. Finally, in the case where only two maxima are detected, one clearly identifies a hand, while the other the head of a visitor. Also in such case it is necessary to go through the remaining hand-points detection phase, to detect either the right or the left hand. In the case where the left hand and the head were detected in the previous step, the right hand is searched as the maximum point in the right direction. The left hand is, instead, searched as the maximum point in the left direction in the opposite situation. As it may be clearer now, the number of points that are searched for in step (e) depends on the number of points that have been initially found with the top-down peaks detection procedure. All the described steps are synthetized in the flow chart shown in Figure 9, while a pictorial representation of the points that are found, as long as the parts that are identified (lower-left, lower-right, top-left and top-right), is provided in the rightmost image of Figure 10. The technical details of our experiment are instead provided in Table 6.

Now, we are interested at analyzing the performance of the described algorithm. Unfortunately, we have not had the chance, yet, of experimenting it in the context of a real museum. However, confident that it may be suitable for such type of scenarios, we implemented and tested it on a number of different frames where a person was moving and assuming different postures.

Figure 10. (Left) Captured frame. (Center) Corresponding blob. (Right) Areas obtained during the proportion extraction step and final detected points (blue = head, green = left hand, yellow = right hand, white = right foot and pink = left foot).

Table 6. Technical details.

Development Environment	Mac OSX Xcode
Hardware	MacBook Pro, 2.66 Ghz, Intel Core i7, 8GB RAM
Software	OpenCV libraries for segmentation procedures

In particular, we tested our algorithm on a set of 57 different frames, representing hence a benchmark for our methodology, where a standing person alternatively moved, stood still with its arms folded and performed actions extending its upper limbs. A body part was classified as "hit" (i.e., detected) when the algorithm returned a point that fell on it (e.g., a relevant point that fell anywhere on a forearm, but not directly on the hand, was not considered valid). The results of our tests are shown in Table 7, where it is immediately clear that in all situations, the head and the two feet were always correctly detected, earning all a perfect score. Not perfect is, instead, the precision with which our algorithm detected the upper limbs, within the frames under analysis. The left hand, in fact, was missed in almost one out of three frames, while the right hand in one out of four. On the grounds of these performance results, we decided to analyze more in depth where our methodology proved to be effective as long as where it failed. However, in doing so, we also observed that our algorithm should be more attentively assessed in those situations where a person performed some type of activity, as those corresponded to the cases where an intangible action recognition service could benefit a digital performance system.

Table 7. Algorithm performance on all cases.

	Hits	**Hit Rate**	**Misses**	**Miss Rate**
Head	57	100%	0	0%
Left hand	41	72%	16	28%
Right hand	43	75%	14	25%
Left foot	57	100%	0	0%
Right Foot	57	100%	0	0%

Therefore we divided the 57 frames into two groups, where the first group included all those frames that were captured while some type of action was being performed, while the second included all the others. In the topmost part of Figure 11 (green background), we listed the 39 frames that belong to the first group: in the first few rows, for example, it is possible to distinguish a person raising both hands to the sky (e.g., movement performed when a weight is being lifted, for example), whereas in the last few ones it is possible to recognize the same person raising its right hand. We ran again our algorithm, but this time only on the pool of 39 aforementioned frames: the results reported in Table 8 reveal that when some type of action was ongoing body parts were detected much more accurately, yielding a miss rate that dropped by half for both hands. Now, why this happened, can be readily explained observing the differences between the frames belonging to the first group, and those that belong to the second (bottom part of Figure 11, red background), where no activity was performed. Almost all frames that are in the second group, in fact, exhibit a blob where the person's arms are folded, close to its torso. The frames that are in the first group instead, in the great majority of cases reveal that the person extended its arms, while performing some type of action.

A visitor performing actions, moving its hands, in front of its body, gives a final interesting situation. Our methodology applied to the frames captured by a frontal webcam would not clearly find the positions of the two hands, since our algorithm would detect no evident maxima in any of the four directions (i.e., top, bottom, left and right). The actions performed in such situation, however, can be reliably detected processing the frames captured

from the overhead position (as in *Tortellino X-Perience*, for example). Two webcams, hence, are sufficient to implement all of the services that are useful within the context of a museum: intangible detection, tracking and action recognition.

Table 8. Algorithm performance when an action is performed.

	Hits	**Hit rate**	**Misses**	**Miss rate**
Head	39	100%	0	0%
Left hand	34	87%	5	13%
Right hand	33	85%	6	15%
Left Foot	39	100%	0	0%
Right Foot	39	100%	0	0%

Finally, we should point out that the methodology that we have here presented has not been envisaged to deal with the possibility that any body parts could be hidden behind obstacles at some point, thus resulting out of sight. This same problem, however, would be shared by all those technologies that are based on a frontal vision of a person (e.g., Microsoft Kinect). This said, our approach could however be extended to deal, to some extent, with such problem. In fact, taking advantage of our system where two webcams are deployed (as shown in Figure 8), we can estimate at all times the position of the barycenter of all people in range, also when standing very close to each other. Hence, based on such information and on the positions where any of the relevant body parts were last detected, it is possible to keep tracking and "guessing" where those points are with a good precision as long as they do not remain hidden for too long. We leave the details of such research direction for future investigations in this field.

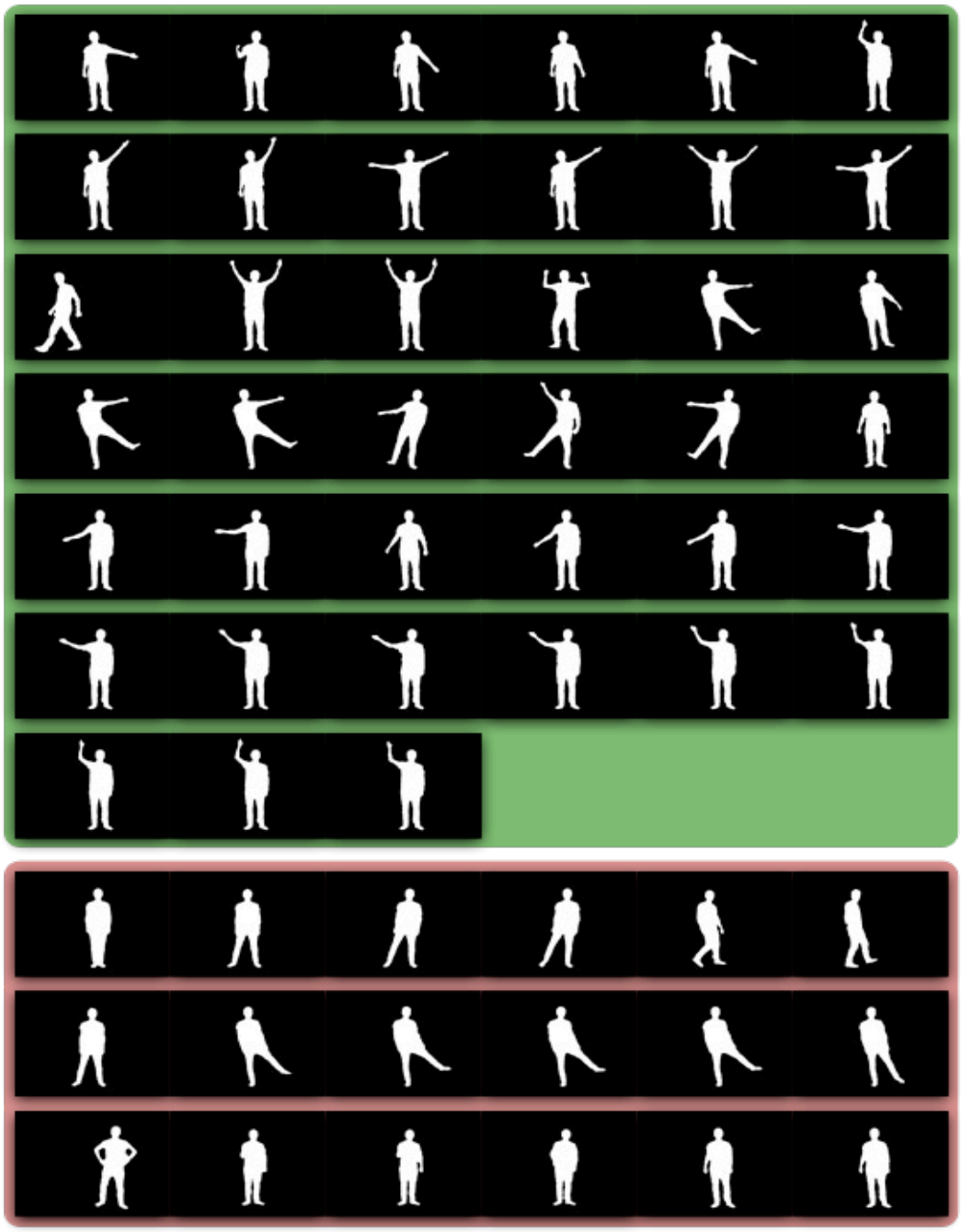

Figure 11. Frames under analysis.

6. Conclusion

Many different and heterogeneous scenarios can be encountered in museums, depending on the particular exhibits that are displayed and on the locations where they are settled. All, however, cannot but prosper from the use of novel and engaging forms of mediation, including those where, for example, visitors perform natural movements and gestures while interacting with their favorite masterpieces. A recent Hollywood blockbuster (which we cite in the title of our paper) provided an extreme example of such vision: in that film visitors are able to interact, physically and verbally, with unanimated collections that suddenly gain life. With this in mind, clearly without claiming of being able to give life to nonliving objects, we here analyzed how gesture and movement-based interactions could be supported with the use of novel technologies. In addition, we also found how such technologies could best serve the purpose of effectively implementing the interactive services that are typically asked for within the context of a museum: (a) visitor detection, (b) visitor tracking, and (c) visitor action recognition. In particular, we demonstrated how all of these services could be supported in a flexible and non-invasive way, not requiring the provision of any device to visitors, but simply using a set of original software algorithms. Finally, we also proved in practice the validity of our approach: two digital performances *Mercator Atlas Robot* and *Tortellino X-Perience*, built following such ideas, have been thoroughly enjoyed by hundreds of visitors.

Acknowledgements

The authors acknowledge the Miur PRIN ALTERNET Project for supporting this work.

References

[1] Cameron, F. and Kenderdine, S., (Eds.) “Theorizing Digital Cultural Heritage: A Critical Discourse”, MIT Press, Cambridge, MA, 2006.

[2] Pasch M., Bianchi-Berthouze N., Van Dijk B. and Nijholt A., “Movement-based Sports Video Games: Investigating Motivation and Gaming Experience”, *Entertainment Computing, Elsevier*, vol. 1, n. 2, pp. 49-61, Apr. 2009.

[3] Pietroni, E., Antinucci, F., “The Approval of the Franciscan Rule. Virtual Experience among the Characters of Giotto’s Work”, in Proc. of Eleventh International Symposium on Virtual Reality, Archeology and Cultural Heritage, Paris, 2010, pp. 171-178.

[4] Rhalibi A., Merabti M., Fergus P. and Yuanyuan S., “Perceptual User Interface as Games Controller”, in *Proc. 5th IEEE Consumer Communications and Networking Conference,* Las Vegas, 2008, pp. 1059-1064.

[5] Vicario L., Monje M., , “Another Guggenheim Effect? The Generation of a Potentially Gentrifiable Neighbourhood in Bilbao”, *Urban Studies*, vol. 40, pp. 2451-2468, 2003.

[6] Carrozzino M. and Bergamasco M., “Beyond Virtual Museums: Experiencing Immersive Virtual Reality in Real Museums”, *Journal of Cultural Heritage*, Elsevier, vol. 11, pp. 452-458, 2010.

[7] Feng-Sheng C., Chih-Ming F., Chung-Lin H., “Hand Gesture Recognition Using a Real-Time Tracking Method and Hidden Markov Models”, *Image and Vision Computing*, vol. 21, no. 8, pp. 745-758, 2003.

[8] Rubegni, E., Memarovic, N., and Langheinrich, M., “CATS: Using Scenario Dramatization to Rapidly Design Public Displays for Stimulating Community Interaction”, in Proc. *of the 29th ACM International Conference on Design of communication* (SIGDOC '11), 2011.

[9] Ghiani, G., Paternò, F., Santoro, C. and Spano, L.D., “UbiCicero: A Location-Aware, Multi-Device Museum Guide”, *Interacting with Computers*, Elsevier, vol. 21, no. 4, pp. 288-303, 2009.

[10] Linge, N., Bates, D., Booth, K., Parsons, D., Heatly, L., Webb, P., Holgate, R., “Realizing the Potential of Multimedia Visitor Guides: Practical Experiences of Developing mi-Guide”, Museum Management and Curatorship, Taylor and Francis, vol. 27, no. 1, pp. 67-82, 2012.

[11] Ciavarella, C. and Paternò, F., “Design of a Handheld, Location-Aware Guide for Indoor Environments”, Personal and Ubiquitous Computing, Springer, vol. 8, n. 2, pp. 82–91, 2004.

[12] Bay H., Fasel B. and Van Gool L., “Interactive Museum Guide: Fast and Robust Recognition of Museum Objects”, in Proc. *of the First International Workshop on Mobile Vision*, Graz, 2006.

[13] Stock O., Zancanaro Z., Busetta P., Callaway C., Kruger A., Kruppa M., Kuflik T., Not W. and Rocchi C., "Adaptive, intelligent presentation of information for the museum visitor in PEACH", *User Modeling and User-Adapted Interaction*, Springer, vol. 17, n. 3, pp. 257-304, 2007.

[14] Papagiannakis G., Singh G. and Magnenat-Thalmann N., "A Survey of Mobile and Wireless Technologies for Augmented Reality Systems", *Comput. Animat. Virtual Worlds*, vol. 19, n. 1, pp. 3-22, 2008.

[15] Roccetti M., Marfia G., Amoroso A., Caraceni S. and Varni A., "Augmenting Augmented Reality with Pairwise Interactions: The Case of Count Luigi Ferdinando Marsili Shooting Game", in *Proc. of the 4th IEEE International Workshop on Digital Entertainment, Networked Virtual Environments, and Creative Technology (DENVECT'12) - 9th IEEE Communications and Networking Conference (CCNC 2012)*, Las Vegas, 2012.

[16] Reis, T., de Sa, M. and Carrico, L., "Multimodal Artefact Manipulation: Evaluation in Real Contexts", in *Proc. of Third International Conference on Pervasive Computing and Applications*, Istanbul, 2008, pp. 570-575.

[17] Arnab S., Petridis P., Dunwell I. and de Freitas S., "Enhacing Learning in Distributed Virtual Worlds through Touch: a Browser-Based Architecture for Haptic Interaction", *Serious Games and Edutainment Applications*, Springer, pp. 149-167, 2011.

[18] Brave S., and Dahley A., "inTouch: a Medium for Haptic Interpersonal Communication", in Proc. *of the CHI '97 extended abstracts on Human factors in computing systems: looking to the future*. Atlanta, 1997, pp. 363-364.

[19] Bergamasco M., "Le Musee del Formes Pures", in Proc. of the 8th IEEE International Workshop on Robot and Human Interaction, Pisa, 1999.

[20] Milosevic B., Farella E. and Benini L., "Continuous Gesture Recognition for Resource Constrained Smart Objects", in *Proc. of the International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*, Florence, 2010.

[21] Zabulis, X., Baltzakis and H., Argyros, A.. "Vision-based Hand Gesture Recognition for Human-Computer Interaction", *The Universal Access Handbook*, Human Factors and Ergonomics, page 34.1 – 34.30. Lawrence Erlbaum Associates, Inc. (LEA), June 2009.

[22] Fujiyoshi, H. and Lipton, A.J., "Real-time Human Motion Analysis by Image Skeletonization", in *Proc. of the Fourth IEEE Workshop on Applications of Computer Vision, 1998. WACV '98.* Proceedings, Princeton, 1998, pp.15-21.

[23] Kehl, R. and Van Gool, L., "Real-time Pointing Gesture Recognition for an Immersive Environment", in *Proc. of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, Southampton, 2004, pp. 577- 582.

[24] Malerczyk, C., "Interactive Museum Exhibit Using Pointing Gesture Recognition", in *Proc. of the 12-th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision,* Plze-Bory, 2004, pp. 165–171.

[25] Manresa C., Varona J., Mas R. and Perales F.J., "Hand Tracking and Gesture Recognition for Human-Computer Interaction", *Electronic Letters on Computer Vision and Image Analysis*, vol. 5, n, 3, pp. 96-104, 2005.

[26] Wang R.Y. and Popovic J., "Real-time Hand-Tracking with a Color Glove", *ACM Trans. Graph.,* vol. 28, n. 3, pp. 1-8, 2009.

[27] Yu, E. and Aggarwal, J.K., "Human Action Recognition with Extremities as Semantic Posture Representation", in Proc. of the *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Miami, 2009, pp.1-8.

[28] Moeslund T.B., Hilton A. and Krüger V., "A Survey of Advances in Vision-based Human Motion Capture and Analysis", *Comput. Vis. Image Underst.*, Elsevier, New York, vol. 104, n. 2-3, pp. 90-126, 2006.

[29] Mitra S. and Acharaya T., "Gesture Recognition: a Survey", *Trans. On Sys., Man and Cyb.*, IEEE, New York, vol. 37, n. 3, pp. 311-324, 2007.

[30] Freeman W. T. Anderson, D.B., Dodge, C.N., Roth, M., Weissman, C.D., Yerazunis, W.S., Kage, H., Kyuma, I., Miyake, Y. and Tanaka, K., "Computer Vision for Interactive Computer Graphics", *IEEE Computer Graphics and Applications*, vol. 18, no. 3, pp. 42-53, 1998.

[31] Snibbe S.S. and Raffle H.S., "Social Immersive Media: Pursuing Best Practices for Multi-User Interactive Camera/Projector Exhibits", in Proc. *of the 27th international conference on Human factors in computing systems*, 2009.

[32] Simpson, Z. "Shadow Garden", in *SIGGRAPH 2002 Electronic Art and Animation Catalog*.

[33] Cohen E., "The Tourist Guide : the Origins, Structure, and Dynamics of a Role", *Annals of Tourism Research*, Elsevier, vol. 12, pp. 5-29, 1985.

[34] Roccetti M., Marfia G. and Zanichelli M., "The Art and Craft of Making the Tortellino: Playing with a Digital Gesture Recognizer for Preparing Pasta Culinary Recipes", *ACM Computers in Entertainment*, ACM, vol. 8, n. 4, 2010.

[35] Sridhar A. and Sowmya A., "Multiple Camera, Multiple Person Tracking with Pointing Gesture Recognition in Immersive Environments", *Advances in Visual Computing*, Lecture Notes in Computer Science, vol. 5358, pp. 508–519, 2008.

[36] Ishi, H., "Tangible Bits: Beyond Pixels," in Proc. *of the ACM Second International Conference on Tangible and Embedded Interaction*, Bonn, 2008.

[37] Marfia G., Roccetti M., Varni A. and Zanichelli M., "Mercator Atlas Robot: Bridging the Gap between Ancient Maps and Modern Travelers with Gestural Mixed Reality", in Proc. *of the 21st IEEE International Conference on Computer Communication Networks (ICCCN 2012) - 8th International Workshop on Networking Issues in Multimedia Entertainment (NIME 2012)*, Munich, July, 2012.

[38] Wren C., Azarbayejani A., Darrell T. and Pentland A., "Pfinder: Real-time Tracking of the Human Body", *IEEE Trans. on Patt. Anal. and Machine Intell.,* vol. 19, no. 7, pp. 780-785, 1997.

[39] Han B., Comaniciu D. and Davis L., "Sequential Kernel Density Approximation through Mode Propagation: Applications to Background Modeling", in *Proc. of the Asian Conference on Computer Vision*, 2004.

[40] Roccetti M. and Marfia G., "Recognizing Intuitive Pre-defined Gestures for Cultural Specific Interactions: An Image-based Approach", in *Proc. of the 3rd IEEE International Workshop on Digital Entertainment, Networked Virtual Environments, and Creative Technology,* Las Vegas, 2011.

[41] Albaum G., "The Likert scale revisited: An alternate version", *Journal of the Market Research Society Market Research Society*, ABI/INFORM Global vol. 39, n. 2, pp. 331, 1997.

[42] Bologna Shanghai World Expo. Available online, accessed on the 20th of June 2011: http://www.bolognaexpo2010.it/zh20/tortellino-x-perience-----.

[43] Giles J., "One per Cent: Video Games Teaches You To Make the Perfect Tortellini", *New Scientist*, Jan. 2011. Available online, accessed on the 20th of June 2011: http://www.newscientist.com/blogs/onepercent/2011/01/com puter-game-that-teaches-you.html.

[44] Mitzman D., "Italian Hi-Tech Software Teaches Perfect Pasta Skills", *BBC News* (06/22/11). Available online, accessed on the 20th of June 2011: http://www.bbc.co.uk/news/world-europe-13856559.

[45] Robertson J., "The Wonderful World of Cooking Interfaces", *Computers, Creativity and Learning (Blog).* Available online, accessed on the 20th of June 2011: http://judyrobertson.typepad.com/judy_robertson/2011/02/the-wonderful-world-of-cooking-interfaces.html