Passive profiling and natural interaction metaphors for personalized multimedia museum experiences

Svebor Karaman, Andrew D. Bagdanov, Gianpaolo D'Amico, Lea Landucci,
Andrea Ferracani, Daniele Pezzatini, and Alberto Del Bimbo

University of Florence,
Media Integration and Communication Center (MICC) Florence, Italy
{svebor.karaman, andrea.ferracani, daniele.pezzatini}@unifi.it,
{bagdanov, damico, delbimbo}@dsi.unifi.it, lea.landucci@gmail.com
http://www.micc.unifi.it/

Abstract. Museums must balance the amount of information given on individual pieces or exhibitions in order to provide sufficient information to aid visitor understanding. At the same time they must avoid cluttering the environment and reducing the enjoyment of the exhibit. Moreover, each visitor has different interests and each might prefer more (or less) information on different artworks depending on their individual profile of interest. Finally, visiting a museum should not be a closed experience but a door opened onto a broader context of related artworks, authors, artistic trends, etc. In this paper we describe the MNEMOSYNE system that attempts to provide such a museum experience. Based on passive observation of visitors, the system builds a profile of the artworks of interest for each visitor. These profiles of interest are then used to personalize content delivery on an interactive table. The natural user interface on the interactive table uses the visitor's profile, a museum content ontology and a recommendation system to personalize the user's exploration of available multimedia content. At the end of their visit, the visitor can take home a personalized summary of their visit on a custom mobile application. In this article we describe each component of our approach as well as the first field trials of our prototype system built and deployed at our permanent exhibition space at *Le Murate*[1] in the city of Florence.

Keywords: Computer vision, video surveillance, cultural heritage, multimedia museum, personalization, natural interaction

1 Introduction

Modern museums are awash in physical and digital content that they struggle to catalog, to maintain, to manage, and – most importantly – to deliver in meaningful ways to the museum-going public. Each visitor would like to easily access different aspects of the massive amount of available information. To
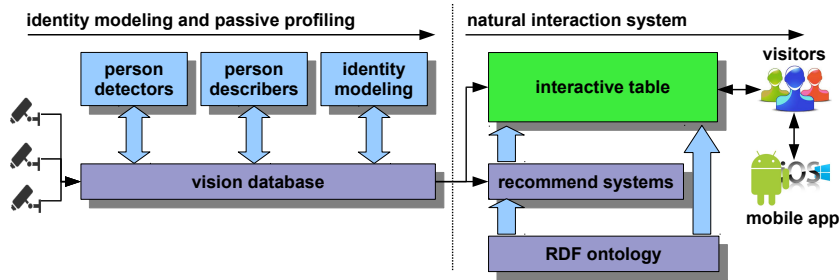
---

[1] http://www.lemurate.comune.fi.it/lemurate/

Fig. 1. An overview of the MNEMOSYNE architecture.

address this issue most research has focused on providing personalized access through handheld devices carried by visitors [1, 5] and possibly offering some sort of augmented reality experience [11, 3]. However, the use of mobile devices is intrusive to the museum experience as it changes the way the visitor behaves with respect to the museum. It also requires active participation of the user in front of each artwork of interest.

Museum exhibits are often designed out of the need to target a sort of "common denominator" visitor. This necessity arises from the difficulty in understanding *a priori* the interests of individual visitors in order to customize content delivery in a meaningful way. User interest modeling for personalization has been adressed in [12] where the user inputs his interests both on the museum website and inside the museum in order to create a "virtuous circle" of online and offline visits. In [8], the authors propose to model user interest based on his displacement in the museum environment in order to personalize audio content delivery via a specific audio guide given to each user. This last approach, like the MNEMOSYNE system we describe in this article, differs from the global trend as it does not require explicit input from the user.

MNEMOSYNE is a three-year research project [2] studying techniques for passively observing museum visitors [9] in order to build profiles of interest for personalizing multimedia content delivery (see figure 1). Aiding the delivery of multimedia content, an ontology that models museum content is used to infer connections between visitor interest and works on display, works located in other collections, and works of broader interest. This knowledge model, along with a statistical recommendation system, is used to drive a natural user interface on a large format interactive table. The user interface allows visitors to explore digital museum content personalized to their own interests, and a mobile application allows them to download content of interest they have explored during their visit to a smartphone or tablet.

In the next section we describe the passive profiling system in detail. In section 3.1 we detail the two types of recommendation systems we have experimented with, and in section 3 we describe the natural interaction user interface used to deliver personalized content. Finally, in section 4 we discuss the first field trials of our prototype system.
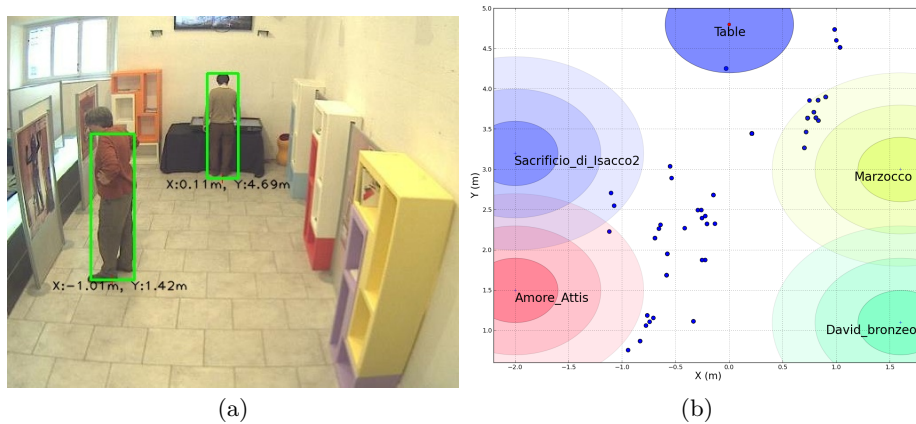
(a)   (b)

Fig. 2. (a) Example of frame with detections (b) Detection map with artwork sphere of influence for one visitor model.

2 Passive interest profiling

We here detail each step of the MNEMOSYNE passive visual profiling approach. First, a pre-processing is required to map the artistic content and the physical properties of the museum. Then, fixed cameras are used to observe visitors as they visit the museum. Our aim is to maintain a record of what they have observed during their visit to build a profile of interest for the visitor.

2.1 Mapping the physical museum

Most museums are often already equipped with a set $\mathcal{C}$ of fixed cameras installed for surveillance purposes and the MNEMOSYNE system is designed to exploit these already-installed cameras. We calibrate each camera c to a common ground plane. A simple tool we allows an operator for estimate the homography $\mathcal{H}_c$ from each camera plane to the ground plane with a few mouse clicks [7].

Given the homography $\mathcal{H}_c$, one can easily input the position on the ground plane of each artwork of interest by simply clicking once in the camera view where the artwork sits on the ground or the position on the floor below the artwork. A sphere of influence is associated with each artwork, defined as a bi-dimensional Gaussian of mean equal to the ground position of the artwork and variance in x and x dimensions defined by the operator. These variances depend on the structure of the museum as well as the artwork scale.

2.2 Identity modeling

On the video stream corresponding to camera $c \in \mathcal{C}$, we run a pedestrian detector [4] in order to obtain a set of N person bounding boxes. The bounding boxes are then described with a number of visual, temporal and spatial descriptors (the

person describers module in figure 1). The descriptor of a person bounding box is defined as:

$$d_i = \left\{ \mathbf{d}_i^a, \mathbf{d}_i^s, d_i^t, d_i^c \right\}, \text{ for } i \in \{1, \ldots, N\}, \tag{1}$$

where $\mathbf{d}_i^a$ is an appearance descriptor consisting of RGB and HS color histograms computed on overlapping horizontal stripes and the HoG (Histogram of Oriented Gradients) descriptor [6] as proposed for person re-identification in [10], $\mathbf{d}_i^s = (d_i^x, d_i^y)$ is the absolute position of the person detection on the ground plane, d_i^t is an integer timestamp, and d_i^c is an index indicating that the detecion comes from camera c. All video streams are synchronized so that d_i^t and d_j^t are comparable.

The fundamental step in passive profiling is associating the detections $D = \{d_i \mid i = 1 \ldots N\}$ to one another to form groups representing individual visitors in the museum. Algorithm 1 details the procedure used to build identity models and to associate detections to them. This algorithm relies on the computation of the distance between a model cluster m_j and a detection description d_i which takes into account the appearance and all spatio-temporal information available. Precisely, the distance between a description d_i and model m_j is computed as:

$$\text{dist}(m_j, d_i) = (1 - \alpha - \beta) \times ||\mathbf{m}_j^a - \mathbf{d}_i^a||_2 \text{ (appearance contribution)} \tag{2}$$
$$+ \alpha \times \text{dist}_w(\mathbf{m}_j^s, \mathbf{d}_i^s, w_s) \quad \text{(spatial contribution)} \tag{3}$$
$$+ \beta \times \text{dist}_w(m_j^t, d_i^t, w_t) \quad \text{(temporal contribution)} \tag{4}$$

where $\text{dist}_w(x, y, w)$ is the windowed L2 distance:

$$\text{dist}_w(x, y, w) = \min(\frac{||x - y||_2}{w}, 1). \tag{5}$$

The parameters w_s and w_t are, respectively, the spatial and temporal window around observations. The weights α and β control the contribution of spatial and temporal distances, respectively, to the overall distance calculation and are defined such that $\alpha, \beta \in [0, 1]$ and $\alpha + \beta < 1$. A detection is associated with a model if its distance to the model is less than δ. The system must accumulate at least τ detections in a temporary model before promoting it to a real one. The appearance of a model $\mathbf{m}_j^a$ is computed as a running average of the detections associated to it, while the position and time information are those of the last matched detection. Whether a model is active determined by the last associated detection time. Note that we also forbid multiple associations from one camera at the same timestamp to the same model.

2.3 Interest profiling

Each visitor's interest profile is built on-the-fly when the visitor enters the interactive table area (see figure 2b) and is sent to the interactive table. Every detection associated with the visitor's model contributes to each artwork according to its proximity to the artwork sphere of influence. If the visitor leaves the interactive area, goes and sees some other artworks and comes back to the table, his interest profile will be updated.

Data: D, δ, τ
Result: Detection associations
$M_a \leftarrow$ getActiveModels()
$M_{temp} \leftarrow$ getTmpModels()
for $d_i \in D$ **do**
 | **dist** $\leftarrow \{\text{dist}(d_i, m_j), \forall m_j \in M_a\}$
 | **if** $\min(\textbf{dist}) \leq \delta$ **and** $M_a \neq \emptyset$ **then**
  | | $k \leftarrow \text{argmin}(\textbf{dist})$;
  | | $m_k.\text{associate}(d_i)$;
 | **else**
  | | **tmpDist** $\leftarrow \{\text{dist}(d_i, m_j), \forall m_j \in M_{temp}\}$
  | | **if** $\min(\textbf{tmpDist}) \leq \delta$ **and** $M_{temp} \neq \emptyset$ **then**
   | | | $k \leftarrow \text{argmin}(\textbf{tmpDist})$;
   | | | $m_k.\text{associate}(d_i)$;
   | | | **if** $m_k.\text{AssociationsCount} \geq \tau$ **then**
    | | | | $M_a = M_a + \{m_k\}$;
    | | | | $M_{temp} = M_{temp} \setminus \{m_k\}$;
   | | | **end**
  | | **else**
   | | | $M_{temp} = M_{temp} + \{d_i\}$;
  | | **end**
 | **end**
end

Algorithm 1: Detection association algorithm.

3 The augmented museum experience

The second stage of the museum visit consists of entering an area equipped with an interactive tabletop display in which users can have a seamless multimedia experience and increase their knowledge of art and cultural heritage. Thanks to the natural interaction with the table, they can explore and select in-depth and related information about each artwork according to suggestions by the recommendation systems, further personalizing the profile of interest built passively during their visit. Their favorite artworks and related multimedia content can then be transferred to a smartphone or tablet via a dedicated mobile application.

3.1 Recommendation systems

The MNEMOSYNE prototype uses two solutions for generating content recommendations to users: a Knowledge-based and an Experience-based system. These Modules have been developed as web servlets which expose several web services accessible via a Representational State Transfer (REST) interface.

Knowledge-based recommendation As a use case we chose one of the most famous museums of Florence, the National Museum of Bargello. In particular, we focus on a subset of 8 monitored artworks from the 70 artworks presented in the *Sala di Donatello*. The MNEMOSYNE Semantic Search Engine exploits the

potential of the Semantic Web through a RDF (Resource Description Framework) that models not only instances of these artworks, but also places, events, historical curiosities and other artworks from Florence and all over the world, and their relations to the 8 chosen artworks. Our ad-hoc ontology implements six different entities: artist, artwork, category, museum/place and story. Every artwork instance is equipped with information about its creator, meanings, materials, techniques and historical context (all represented by instances of their own). Each artwork entity is also accompanied by a variety of multimedia content and has different thematic links to other artworks or stories. SPARQL queries are used to query subgraphs of data from the ontology, providing different views on the data model: subgraphs of artwork data, stories related to artworks, and resources related by tags or stories.

Experience-based recommendation The MNEMOSYNE Recommendation Engine uses two types of recommendation algorithms. Recommendation algorithms try to solve a problem of prediction, they give the system the ability to suggest to users items that they are more likely to find interesting. The Recommender Engine implements metrics based on both user similarity and item similarity. The data model consists of preferences stored as triples in a database table containing the following fields: the user ID, the item ID, and a value, assigned by the MNEMOSYNE passive visual profiling module, expressing the strength of the user preference for the item. From this information we compute which users or items are more similar. Both similarity metrics, user based and item based, make use of the same components: a data model, a metric of similarity, a notion of proximity (i.e. a neighborhood of users or items) and an algorithm that predicts values of preference weighting them differently according to the similarity metric. The Recommendation Engine uses Euclidean distances: a greater distance indicates a lower similarity. The system makes use of the Mahout library[2], the state of the art for machine learning on big data.

3.2 The natural interaction tabletop system

When one visitor approaches the table, he is detected by the camera which triggers the tabletop system and the computer vision system to exchange data about his profile of interest that had been built during the previous experience with the physical artworks. Initially, the display of the tabletop shows a language selector and then a quick interactive tutorial in the form of a carousel of images.

The metaphor used for the user interface is based on the idea of the hidden museum waiting to be unveiled, an all-digital environment where visitors can go beyond the artworks level of the physical museum. The tabletop provides an augmented perspective of the museum, thanks to the interest profile estimated by the computer vision system, the recommendation systems and via the adoption of natural interaction paradigms. With the natural interaction interface, visitors can discover new, interesting information and resources, and then collect them for future inspection at home. The main user interface adopts common principles
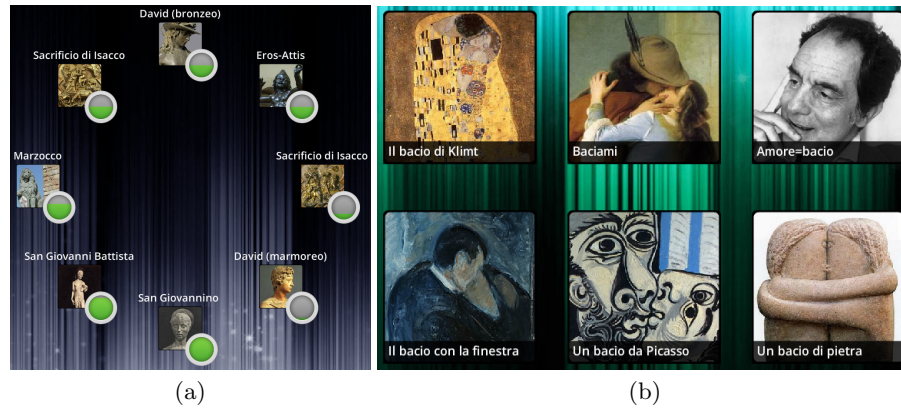
---

[2] http://mahout.apache.org

Fig. 3. (a) Detail of the artwork level: the artworks of the museum are represented with the original title, a thumbnail and a circular symbol visualizing the amount of interest showed by the current user during their visit. (b) Detail of the insights space in the related resources level: information related thematically to the selected artwork.

derived from Information Visualization and consists of two different levels of navigation: the *artworks* and the *related resources*.

The artworks level visualizes artworks of the museum for which the visitor has shown the highest level of interest, see figure 3a, based on the profile created by the computer vision system. When the user touches an artwork an animation starts in order to trigger the following actions: the space containing the unselected artworks moves up vertically off the screen, the selected artwork moves to the middle left and a new space of information comes up from the bottom.

The related resources level follows the selection of an artwork and consists of a structured environment in which the visitor can navigate the multimedia content related to this artwork. Related resources are organized in four different spaces (*stories*, *secrets*, *recommendations*, *insights*) according to their relations to the artwork based on the recommendation system and the ontology structure:

- **stories:** stories directly related to the artwork in the ontology;
- **secrets:** resources related to the artwork and its related stories in the ontology according to the knowledge-based recommendation system;
- **recommendations:** similar artworks according to the experience recommendation system using the visitor profile;
- **insights:** resources related by tags or stories to the artwork according to the knowledge-based recommendation system.

The user can navigate these spaces via a textual menu on the bottom of the display. In each space details are represented in a different layout: the background color, shape and arrangement of interactive items are all adjusted in order to highlight the difference between the visualized data and to provide the user a pleasant experience while navigating in the information space. The *insights*
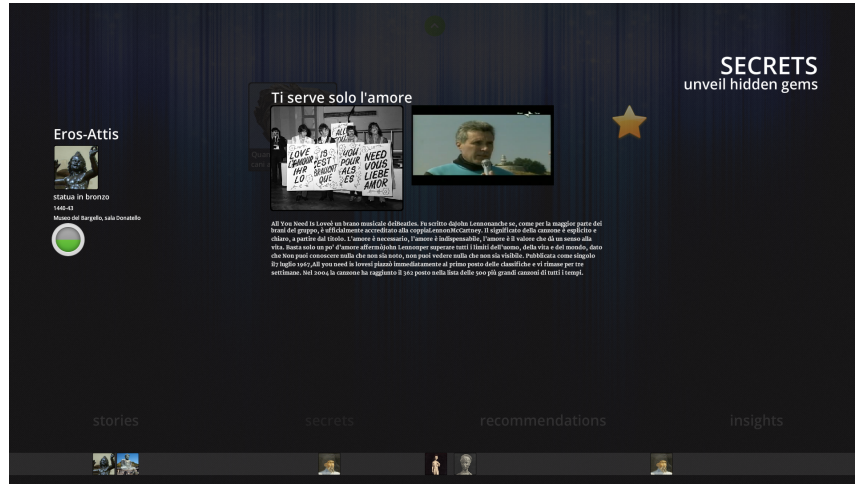
Fig. 4. The information level: when the user selects an item from one space, multimedia content is displayed in the center of the screen. A *star* element can be touched in order to collect the item in the space at the bottom of the screen. The selected artwork is visualized on the left of the screen with its title, a thumbnail, location and level of visitor interest.

space is illustrated in figure 3b. In each space, if the user touches one of the visible items it triggers an information window composed of several multimedia contents, see figure 4. The visitor is able to collect the most interesting resources for later use via the mobile system.

3.3 Mobile system

The main goal of the MNEMOSYNE mobile application, shown in figure 5, is to collect personalized digital content displayed in the interactive tabletop interface. Unlike [5, 11], the mobile app is intended to be used at the end of the visit and not as an interactive device throughout the museum tour. It was developed using the Adobe AIR framework and can therefore be installed on devices running iOS (iPhone, iPad) or Android.

Once installed on the visitor's smartphone, the application allows the user to scan a QR code displayed on the interface of the interactive table. This transfers a unique content identifier user's mobile device. The application queries the MNEMOSYNE database to retrieve the user's favorite artworks, generated both through the passive profiling module and from their interaction on the tabletop surface. The user then has access to in-depth information about individual and related artworks or resources in the MNEMOSYNE dataset. In particular, the user can visualize a collection of points on a map of interest suggested by the recommendation system taking in account the user's profile and current geolocalization. The latter functionality allow us to extend the personalized user experience of the visit from an indoor to an outdoor scenario.
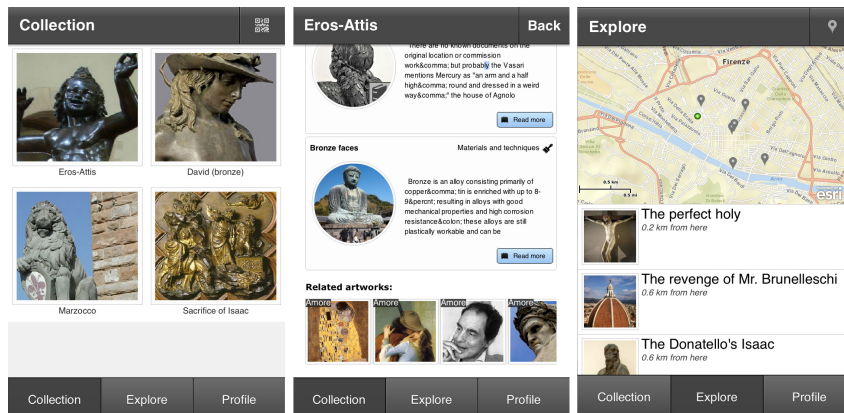
Fig. 5. The mobile application: user's favourites artworks (on the left), in-depth of the artwork (on the center) and the map of suggested points of interest(on the right).

4 Field trials and ongoing work

We have installed a prototype of the MNEMOSYNE system in a permanent exhibition space at *Le Murate* in the historical center of Florence. For this installation, we printed four high resolution images of artworks from the Donatello Room and used a single surveillance camera to capture images of visitors observing the artworks as well as interacting with the table. Some detections obtained are shown in figure 2a. For this installation the parameters were set manually ($\alpha = \beta = 0.2$, $w_s = 5$m, $w_t = 80$ frames, $\delta = 0.75$ and $\tau = 10$), but given a training set of annotated detections they could be easily learned. Two critical issues were evident when running the system continuously for several hours: ensuring that the system does not lag and that profile messages are thus sent in a timely fashion; and limiting the confusion between visitors since, when observing for several hours, it is very likely that several persons will have similar appearance.

Lag is mostly due to the detection process which is computationally onerous, and we dedicated an 8-core computer to this task and limited the frame rate to 5 frames per second. Moreover, we implemented a lag monitor that considers a maximum allowed lag (set to 5 seconds) and discards frames until the lag falls within the allowed range. To limit confusion between visitors with similar appearance across several hours of observation we limit the association of detections to visitor models that were "active", i.e. ones with which at least one detection was associated in the previous 10 minutes. These proposed solutions are the first necessary steps towards bringing MNEMOSYNE to real museum.

5 Conclusions

In this paper we detailed our proposal to enhance and personalize museum visits. Our system makes use of passive observation to estimate each visitor's interest.

This interest profile is then used conjointly with a recommendation system to provide personalized content delivery through a natural interaction interface. MNEMOSYNE is operational and has been tested, but in order to determine the efficiency of the approach several aspects must be evaluated with further experiments, user studies, and also potentially on how the suggested resources impact visitor visits. We are also interested in using the mobile application as a bridge between different museums adopting MNEMOSYNE, by suggesting artworks of interest in other museums. These are our main objectives and we hope to deploy our system in major museum in Florence in next few months.

References

1. Baber, C., Bristow, H., Cheng, S.L., Hedley, A., Kuriyama, Y., Lien, M., Pollard, J., Sorrell, P.: Augmenting museums and art galleries. In: Human-Computer Interaction. In: INTERACT. vol. 1, pp. 439–447 (2001)
2. Bagdanov, A.D., Del Bimbo, A., Landucci, L., Pernici, F.: Mnemosyne: Enhancing the museum experience through interactive media and visual profiling. In: Multimedia for Cultural Heritage, pp. 39–50. Springer (2011)
3. Bay, H., Fasel, B., Van Gool, L.: Interactive museum guide: Fast and robust recognition of museum objects. In: Proceedings of the first international workshop on mobile vision (May 2006)
4. Bimbo, A.D., Lisanti, G., Masi, I., Pernici, F.: Person detection using temporal and geometric context with a pan tilt zoom camera. In: Pattern Recognition (ICPR), 2010 20th International Conference on. pp. 3886–3889. IEEE (2010)
5. Bruns, E., Brombach, B., Zeidler, T., Bimber, O.: Enabling mobile phones to support large-scale museum guidance. MultiMedia, IEEE 14(2), 16–25 (2007)
6. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. vol. 1, pp. 886–893. IEEE (2005)
7. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, ISBN: 0521540518, second edn. (2004)
8. Hatala, M., Wakkary, R.: User modeling and semantic technologies in support of a tangible interface. Journal of User Modeling and User Adapted Interaction 15(3-4), 339–380 (2005)
9. Karaman, S., Bagdanov, A.D.: Identity inference: generalizing person re-identification scenarios. In: Computer Vision–ECCV 2012. Workshops and Demonstrations. pp. 443–452. Springer (2012)
10. Karaman, S., Lisanti, G., Bagdanov, A.D., Del Bimbo, A.: From re-identification to identity inference: labelling consistency by local similarity constraints. In: Person Re-identification. To appear (2013)
11. Kuflik, T., Stock, O., Zancanaro, M., Gorfinkel, A., Jbara, S., Kats, S., Sheidin, J., Kashtan, N.: A visitor's guide in an active museum: Presentations, communications, and reflection. Journal on Computing and Cultural Heritage (JOCCH) 3(3), 11 (2011)
12. Wang, Y., Stash, N., Sambeek, R., Schuurmans, Y., Aroyo, L., Schreiber, G., Gorgels, P.: Cultivating personalized museum tours online and on-site. Interdisciplinary Science Reviews 34(2-3), 2–3 (2009)