This is a preprint version, the final publication is available at Springer via:
http://dx.doi.org/10.1007/s11042-014-2192-y
Multimed Tools Appl

Personalized multimedia content delivery on an interactive table by passive observation of museum visitors

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

Received: 14 December 2013 / Revised: 25 June 2014 / Accepted: 13 July 2014

Abstract The amount of multimedia data collected in museum databases is growing fast, while the capacity of museums to display information to visitors is acutely limited by physical space. Museums must seek the perfect balance of information given on individual pieces in order to provide sufficient information to aid visitor understanding while maintaining sparse usage of the walls and guaranteeing high appreciation of the exhibit. Moreover, museums often target the interests of average visitors instead of the entire spectrum of different interests each individual visitor might have. Finally, visiting a museum should not be an experience contained in the physical space of the museum 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 address these issues through a new multimedia museum experience. Based on passive observation, the system builds a profile of the artworks of interest for each visitor. These profiles of interest are then used to drive an interactive table that personalizes multimedia content delivery. The natural user interface on the interactive table uses the visitor's profile, an ontology of museum content and a recommendation system to personalize exploration of 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 in detail 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 Florence together with the first results of the evaluation process during the official installation in the *National Museum of Bargello* [2].

Svebor Karaman · Lea Landucci · Gianpaolo D'Amico · Andrea Ferracani · Daniele Pezzatini · Alberto Del Bimbo
Media Integration and Communication Center (MICC)
University of Florence, Italy
E-mail: firstname.lastname@unifi.it

Andrew D. Bagdanov
Computer Vision Center, Barcelona
E-mail: bagdanov@cvc.uab.es

[1] http://www.lemurate.comune.fi.it/lemurate/
[2] http://www.uffizi.firenze.it/musei/?m=bargello

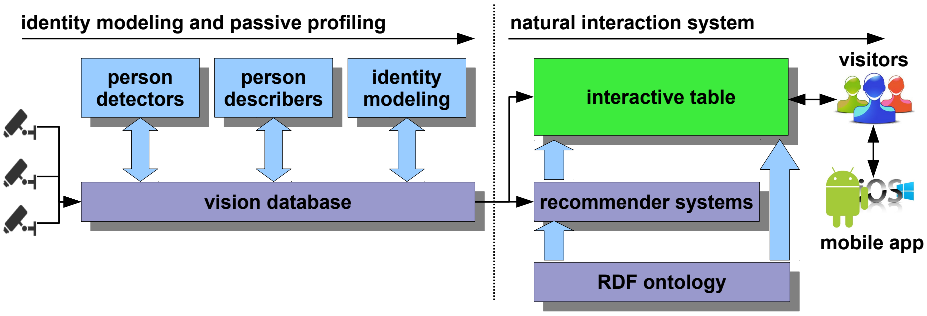


Fig. 1 An overview of the MNEMOSYNE system architecture.

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

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's interests and knowledge induces a unique perspective of what is relevant among the massive amount of available information. Determining what is most relevant and how to deliver this content to the user has been the focus of several works in the literature. To address this issue most research has focused on providing personalized access through handheld devices carried by visitors [3, 17] and possibly offering some sort of augmented reality experience [9, 33]. 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. However, as stated in [14], personalization enables changing "the museum monologue" into "a user-centred information dialog" between the museum and its visitors. Customizing content delivery in a meaningful way has been the subject of research of a few works. User interest modeling for personalization has been addressed in [56] 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 [25], 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[3] system we describe in this article, differs from the global trend as it does not require explicit input from the user.

[3] Mnemosyne (the source of the word *mnemonic*) was a titan and the personification of memory in Greek mythology. We chose the name MNEMOSYNE for the project to emphasize the assistive nature of the proposed system in acquiring new information.

MNEMOSYNE is a three-year research project [5] studying techniques for passively observing museum visitors [30] in order to build profiles of interest for personalizing multimedia content delivery (see figure 1). Aiding the delivery of multimedia content, an ontology modeling all available multimedia content and their relationships is used to infer the most suitable content based on the estimated visitor interests. The ontology enables connection with works on display, but also with works located in other collections and even with 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 discuss work from the literature related to the goals of the MNEMOSYNE project. We then describe the passive profiling system in detail in section 3. In section 4.2 we detail the two types of recommendation systems we have experimented with, and we describe the user interface of the interactive table used to deliver personalized content. Finally, in section 5 we report on the experimental validation of our profiling approach, user satisfaction surveys, and the ongoing field trials of our prototype system.

2 Related work

In this section we describe elements of the state-of-the-art most relevant to the goals of the MNEMOSYNE project: Natural Interaction, Content Personalization and Computer Vision. We begin with an overview of interactive multimedia museums in terms of motivation, development and natural interaction solutions. Content Personalization is discussed in section 2.2. In section 2.3 we discuss the state-of-the-art in Computer Vision, with particular attention to the key technologies employed in the MNEMOSYNE prototype. Finally, in section 2.4 we contextualize the MNEMOSYNE system with respect to the state-of-the-art with a discussion of the technological choices made in developing the prototype.

2.1 Interactive multimedia museums

Technological advances over the last decade have transformed the museum experience, making it more personalized, intensive and engaging both at a virtual (on-line) and physical (on-site) level. The purpose of museums is shifting from providing static information to providing personalized services to a broad range of visitors worldwide. Personalized content delivery also enables changing "the museum monologue" towards a user-centred dialog between the museum and its visitors [14]. These advances are perhaps best exemplified by the accelerating development of online offerings from museums and the changing nature of information access due to the proliferation of ubiquitous, low-cost mobile devices [1].

Museum tours offer visitors a unique museum experience providing many insights on the artworks of an exhibit. There are four principal types of museum tour: human-guided tours, audio tours, online/virtual tours, and multimedia tours [55].

The traditional human-guided and audio tours are still available in most museums, but in recent years Web and mobile technologies have dramatically increased access to museum content. Here we briefly review the first experiments with online and mobile museum tours.

Online/virtual tours: One of the earliest examples of a virtual museum was WebLouvre which delivered high-quality (for the time) copies of works of art by leading historical artists [13]. Following this ground-breaking example, galleries around the world have initiated projects to build networks for on-line fruition of artwork and cultural heritage resources [57,58,64,65]. By 2014, all major museums in the world offer a complete and engaging website to visitors for planning their tours and deepening their knowledge about artworks on exhibit before or after their visit [60–63].

Mobile multimedia tours: More and more museums are offering multimedia tours implemented on a range of mobile devices. These tours allow visitors more informed enjoyment and hence greater engagement with the artworks [46]. The challenge is now becoming how to deliver a museum experience to visitors in an immersive museum environment [56]. Interactivity and instruments for analyzing how people move around museum spaces can transform visitors into active participants in their museum experience [50]. Early experiments in this direction used cumbersome and intrusive tools such as wearable computers [49, 51]. The PEACH (Personal Experience with Active Cultural Heritage) [59] project aims to provide interactive appreciation of cultural heritage by means of accurate reconstructions of objects. The project focuses on natural interaction assisted by microsensory systems, encompassing natural language processing, perception, image understanding, intelligent systems and others.

2.2 Personalization via recommendation

Negroponte, in the chapter on his thoughts about The Post-Information Age in his widely acclaimed book *Being Digital* [38], noted that:

> "True personalization is now upon us. It's not just a matter of selecting relish over mustard once. The post-information age is about acquaintance over time: machines' understanding individuals with the same degree of subtlety (or more than) we can expect from other human beings, including idiosyncrasies (like always wearing a blue-striped shirt) and totally random events, good and bad, in the unfolding narrative of our lives [...]"

It can be difficult for people to find the right information at the right time and at the right level of detail. The literature on personalization and recommendation systems is vast and complex because both themes are central to many disciplines. Researchers have developed adaptive systems that adjust their behavior to the goals, tasks and interests of individuals and groups. Such systems differ from static solutions in the use of a user model representing user characteristics and in the dynamic creation of content and presentations adapted to each user [18].

In museum contexts, personalization is a communication strategy based on collaboration, learning and adaptation between the museum and its visitors [14]. Multimedia guides should support personalization of information delivered and should enable an experience adapted to each visitor's own pace and interests. To achieve this, information must be presented in a manner that is appropriate to the physical location of the visitor as well as to the location of the works of art within the environment. By connecting information in an exhibit and presenting it to the visitor in a coherent way with respect to his physical location, the overall experience is optimized [52]. Personalization in this sense is related to the concept of situation-aware content, where information is most effective if presented in a cohesive way built on previously delivered information. This can be accomplished using references to space and time, which aids the visitor in orienting himself with respect to available works of art [67].

Basic approaches to personalization: We consider personalized systems aimed at learning user preferences and providing recommendations automatically [41]. Approaches to personalization can be broadly categorized as:

- **Collaboration-based methods** which identify peers of a visitor having similar known preferences and recommend those items that were most liked by peers. Collaboration-based methods, however, suffer from the sparsity problem: data reflecting user preferences is usually sparse and insufficient to identify similarities. Huang et al. [26] propose an associative retrieval framework to address this problem.
- **Content-based methods** analyze common features among items a visitor liked and recommend items that are similar. Acquiring the preferences of a user is the bottleneck for content-based methods, and it can be unclear how similarity between content should be measured. The diversity of recommendations is often a desirable feature in content-based recommender systems [11].
- **Model-based algorithms** use the collection of visitor ratings to learn a model then used to make predictions. The authors of [34] used a simple probabilistic model to demonstrate that collaborative filtering can be valuable, even when relatively little data for each user is available. Visitors are usually clustered into groups of similar interests to facilitate group-based personalization (see for example [48, 53]).
- **Hybrid approaches** combine collaborative and content-based methods in order to overcome the limitations of each approach. The three most common ways of combining collaborative and content-based methods are: (i) combining recommendations with voting or selection mechanisms [21]; (ii) introducing elements of one recommendation type into another, for example by reducing the dimensionality and sparseness of the ratings matrix [21]; and (iii) building generic models that include elements of both, such as probabilistic latent semantic analysis [42]).

Non-intrusiveness: Minimization of feedback requests is essential in personalization systems. Completely non-intrusive feedback determination methods have been proposed in the literature, but such techniques are often inaccurate. The problem is thus usually reformulated to minimize intrusiveness while learning visitor preferences accurately. Recently, techniques exploiting item popularity (well- vs. poorly-known artworks) and item controversy (similarly vs. differently rated artworks) have been applied to the visitor interest learning problem [44, 47]. Non-intrusiveness is still as a major obstacle to advancing the state-of-the-art in recommender systems [48].

2.3 Computer vision for profiling

The MNEMOSYNE project uses an array of computer vision techniques to perform visual profiling of visitor interest at a distance. In potentially crowded museum interiors, the most important objective of the visual profiling system is to accurately identify visitors moving about in the museum environment. Person detection is a technique that has advanced dramatically in recent years, mostly due to interest from the autonomous vehicle community [20, 36, 37, 40, 66]. Some of the most successful techniques for person detection use sophisticated body-part models [23] and explicit occlusion modeling [54]. Most state-of-the-art detection techniques, however, do not satisfy the strict real-time requirements of the MNEMOSYNE application scenario.

A traditional approach would prescribe person tracking after detection. A modern trend in visual tracking is the tracking-by-detection approach [15, 35]. In this approach, detectors are used to drive optimal estimation of a sequence of position estimates in a joint optimization of detector and tracker outputs. These approaches are particularly attractive in static, indoor environments where strong environmental priors can be learned in order to predict where interesting targets are likely to appear, and where they are likely to go [16].

For visual profiling applications, trackers must provide reliable tracks over very long sequences. In such cases, a more reliable alternative can be Person Re-identification instead of tracker [32]. Person re-identification is the problem of identifying previously seen individuals on the basis of one or more images captured from one or more cameras. Person re-identification is usually performed using a combination of sophisticated visual descriptions of person images [10, 19, 22] and learning methods to accurately identify previously-seen persons [2, 31, 32, 43].

2.4 Discussion

To be truly transformative, the technology used for all aspects of visitor profiling and customized content delivery must become transparent to the user [39]. Unobtrusiveness was an overarching design goal in the MNEMOSYNE project, and as such this critical factor makes a transversal appearance in the motivation for each principal aspect of the MNEMOSYNE technological platform.

Computer vision technologies for profiling have multiple advantages: (i) they are seamless and non intrusive in that they are unseen by visitors and can often piggyback on existing camera systems used for video surveillance; (ii) they are scalable in that the size of deployment can be personal or very large; and (iii) they are evolutive and future proof as they rely on cameras and new features usually means software upgrade only. Due to these considerations in the MNEMOSYNE project we adopted a “hands off” visual profiling approach which uses exclusively cameras and computer vision techniques for visitor interest profiling. The main bottleneck in our visual profiling system is the detection and person description process. We use an adaptive state-of-the-art detector [8] and a person descriptor [32], both developed in our laboratory, that allow the visual analysis to be performed in real-time. Our approach to visual profiling is reminiscent of both tracking and person re-identification. Indeed, the system should be able to retrieve all positions occupied in all cameras by each visitor during his visit. Multiple target tracking, however, is generally neither reliable enough nor feasible in real-time in crowded multi-camera museum scenarios. Hence, while our single camera profiling approach is closer to tracking, the method of obtaining the complete profile of interest from all cameras is closer to re-identification. In the next section we describe the computer vision system used in MNEMOSYNE for profiling visitor interest.

Rather than rely on each visitor's mobile device for content delivery, as is common with most mobile multimedia tours, in the MNEMOSYNE project we believe it is better to concentrate information delivery on dedicated, natural interaction displays in the museum environment. We feel that the use of mobile devices distracts the visitor from artworks on display. Rather than oblige the visitors to be equipped with their own mobile device, users interact with a user interface personalized according to their profile of interest. This user interface uses museum information content organized in the form of an ontology in the Resource Description Framework (RDF) [27], guaranteeing that the information resources are extensible and that external resources can be easily and liberally added to the ontology of cultural heritage resources. We discuss the natural interaction user interface used in the MNEMOSYNE prototype in section 4.2.

Content recommendations should ideally be made on the basis of the visitor's profile of interest and the history of his interactions with the available content. Requiring the user to manually select interests or navigate hierarchies of topics in the user interface goes against our vision of unobtrusiveness for the MNEMOSYNE system. In the MNEMOSYNE prototype we take a hybrid approach to recommendation that leverages both knowledge-based and experience-based recommendation to personalize content delivery (see section 4.1).

3 Passive interest profiling

Here we detail each step of the MNEMOSYNE passive visual profiling approach. First, a pre-processing step 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 each visitor has observed during his visit, to build a profile of interest for each visitor separately, and then to use this as the seed for personalized multimedia content delivery.

3.1 Mapping the museum content

As a starting point, we chose to use one of the most famous museums of Florence, the National Museum of Bargello, and in particular the Hall of Donatello. Inside this salon there are more than 70 artworks, the majority of which by Donatello. We focused on a subset of ten artworks which are illustrated in Fig. 2.

One of our main design principles was to create an ontology containing not only instances of the artworks shown in Fig. 2, but also places, events, historical curiosities and artworks contained in other museums in Florence and all over the world, all related in some way to the ten chosen works. Such an ontology can

Fig. 2 The ten artworks modeled in the MNEMOSYNE ontology: (a) *Attis* by Donatello, (b) *David* (bronze) by Donatello, (c) *Madonna con bambino* by Michelozzo, (d) *David* (Marble) by Donatello, (e) *Marzocco* by Donatello, (f) *San Giorgio* by Donatello, (g) *Madonna con la mela* by Luca della Robbia, (h) *David* (Bronze) by Verrocchio, (i) *Sacrificio di Isacco* by Ghiberti and Brunelleschi and (j) *San Giovanni Battista* by Donatello and Desiderio da Settignano.

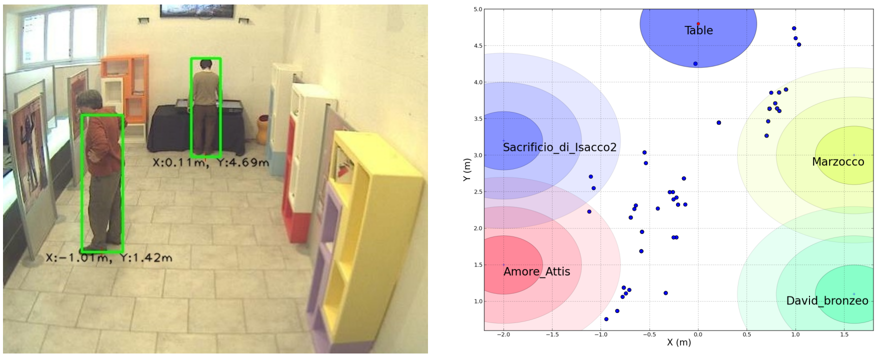
be exploited by the recommender module (see section 4.1) in order to provide customized, profile-based insights to each visitor.

We built an ad-hoc ontology in which we describe and create relations between: the artworks, the artists, the museums, and stories. Indeed, every instance of an artwork is detailed with information about its author, its location, and some related stories. The stories can be elements giving: a description, an interpretation, some historical context, some explanation on the materials, a curiosity, or details of an artistic competition. All stories are accompanied by multimedia content such as images and video. Moreover, each artwork may have different thematic links to other artworks or stories.

We built the ontology using the Resource Description Framework (RDF) language that focuses on the description of digital resources. RDF is now widely used and is an efficient tool applied to various domains of knowledge [27]. The ontology is easily expandable with other resources in order to infer richer results and/or to link to wider contexts.

3.2 Mapping the physical museum

Most museums are 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. Each camera $c \in \mathcal{C}$ should be calibrated to a ground



(a) (b)

Fig. 3 (a) Example frame with detections where the ground floor positions are estimated with the homography $\mathcal{H}_c$. (b) Detection map for one visitor model with artwork spheres of influence and the interactive table area.

plane coordinate system common to all cameras. A simple tool we developed allows an operator to estimate the homography $\mathcal{H}_c$ from each camera plane to the ground plane with a few mouse clicks [24], making this step of the mapping of the physical museum easy and fast.

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 it. If the layout of the museum and the artwork positions are already recorded by the museum staff they can be directly inserted in the MNEMOSYNE backend database. A sphere of influence is associated with each artwork, defined as a bi-dimensional Gaussian with mean equal to the ground position of the artwork and variance in the x and y dimensions. These variances are defined by an operator and represent the area where visitors will tend to stand to observe this artwork. They depend on the structure of the museum as well as the artwork scale. Specifically, to each artwork a_i is associated its ground plane position $\mathbf{a}_i^s = (a_i^x, a_i^y)$ and its influence variance $\sigma_{\mathbf{a_i}} = (\sigma_{a_i}^x, \sigma_{a_i}^y)$. See figure 3b for an example of 4 artworks with their respective spheres of influence plotted as three ellipses representing one, two and three times their respective variances.

3.3 Identity modeling

On the video stream corresponding to camera $c \in \mathcal{C}$, we run a pedestrian detector [8, 12] in order to obtain a set of N person bounding boxes (some example detections are shown in figure 3a). 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 a feature vector describing the appearance of a person. This descriptor consists of RGB and HS color histograms computed on overlapping horizontal stripes and a HOG (Histogram of Oriented Gradients) descriptor [20] as proposed for person re-identification in [32]. The spatial component $\mathbf{d}_i^s = (d_i^x, d_i^y)$ 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. This problem is closely related to person re-identification [29], but real-time constraints, needed to ensure that the interactive table is updated on time to personalize multimedia content delivery, exclude the use of most standard re-identification methods. Moreover, the system must be able to model the entrance and exit of persons in the observed area while the re-identification problem considers that the whole set of identities is known *a priori*. 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:

$$\mathrm{dist}(m_j, d_i) = (1 - \alpha - \beta) \times ||\mathbf{m}_j^a - \mathbf{d}_i^a||_2 \quad \text{(appearance contribution)} \tag{2}$$

$$+ \alpha \times \mathrm{dist}_w(\mathbf{m}_j^s, \mathbf{d}_i^s, w_s) \quad \text{(spatial contribution)} \tag{3}$$

$$+ \quad \beta \times \mathrm{dist}_w(m_j^t, d_i^t, w_t) \quad \text{(temporal contribution)} \tag{4}$$

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

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

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 considered active is 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.

3.4 Interest profiling

Each visitor's interest profile is built on-the-fly when the visitor enters the interactive table area (see the blue "Table" area in figure 3b). We denote the visitor as v and his associated model obtained by Algorithm 1 as m_v. To build the interest profile, we rely on the whole set of detections associated with this person model m_v. Every detection associated with the visitor's model contributes to each

```
Data: D, δ, τ
Result: Detection associations
M_a ← getActiveModels()
M_temp ← getTmpModels()
for d_i ∈ D do
    dist ← {dist(d_i, m_j), ∀m_j ∈ M_a}
    if min(dist) ≤ δ and M_a ≠ ∅ then
        k ← argmin(dist);
        m_k.associate(d_i);
    else
        tmpDist ← {dist(d_i, m_j), ∀m_j ∈ M_temp}
        if min(tmpDist) ≤ δ and M_temp ≠ ∅ then
            k ← argmin(tmpDist);
            m_k.associate(d_i);
            if m_k.AssociationsCount ≥ τ then
                M_a = M_a + {m_k};
                M_temp = M_temp \ {m_k};
            end
        else
            M_temp = M_temp + {d_i};
        end
    end
end
```

Algorithm 1: The detection association algorithm used for visual profiling.

artwork according to its proximity to the artwork sphere of influence. Specifically, the interest $I_v(a_i)$ of visitor v for artwork a_i is estimated as:

$$I_v(a_i) = \sum_{d_i \in \mathcal{D}_v} \exp\left(-\frac{1}{2}\left(\left(\frac{d_i^x - a_i^x}{\sigma_{a_i}^x}\right)^2 + \left(\frac{d_i^y - a_i^y}{\sigma_{a_i}^y}\right)^2\right)\right) \tag{6}$$

where $\mathcal{D}_v$ is the set of detections associated to model m_v.

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. Note that the interest profile is normalized to sum to one and hence represents the distribution of interests of the visitor. Once the profile is computed, it is sent to the interactive table in order to personalize the multimedia content delivered to the user to best match its interests. A visualization of an example profile of interest obtained in our installation at *Le Murate* is given in figure 4, where the length of each color bar correspond to the amount of interest in the corresponding artwork. Note that this profile correspond to the set of detections depicted in figure 3b.

This profiling approach is adequate for an installation where a single camera is used, as was the case for our installation at *Le Murate*. However, for the ongoing installation at the *National Museum of the Bargello*, multiple cameras are mandatory to cover the wider observed area. Indeed, four cameras are necessary to be able to observe visitors near all ten selected artworks.

To extend profiling to multiple cameras, the approach described above is run independently on each camera video stream. When a visitor enters the interactive area, the profiling system queries for similar appearance models from the other cameras and builds the final, global interest profile as a weighted sum of local (i.e. from a single camera) profiles of interest. This procedure is very efficient and enables creating and sending the profile of interest of the visitor approaching the

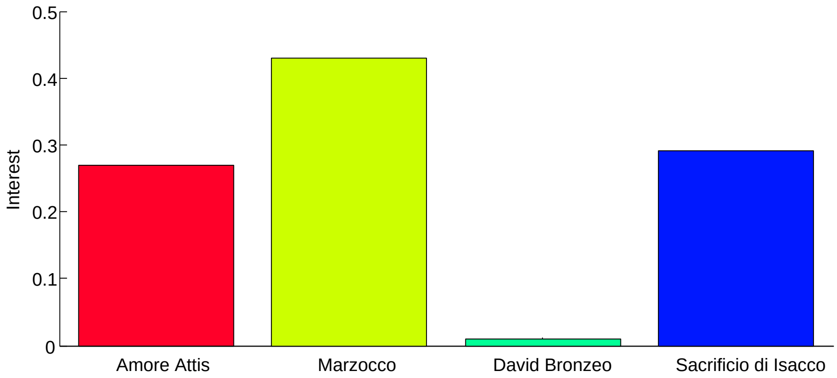


Fig. 4 Example interest profile corresponding to the set of detections depicted in figure 3b

table to the interactive table in time to personalize the interface to the visitor's interests.

4 The augmented museum experience

The second phase of a visitor's visit in the museum consists of a multimedia-driven experience using advanced interactive systems. The exhibit has a dedicated area at the end of the exhibition hall where people can live a second stage of their cultural experience. An interactive tabletop is in this area that allows users to interact with all available multimedia assets which provide in-depth information about the artworks of the museum and beyond. The personalized browsing of multimedia content relies on a recommendation engine connected to the passive profiling system. In addition to the previously described passive profile of interest, a personalized active profile of interest is implicitly built by each visitor while interacting with the tabletop. The combined interest profile can be transferred to a smartphone (or tablet) via a dedicated mobile application, enabling the ability to suggest interesting places to see in Florence and beyond. In this section we discuss the recommendation system, the tabletop system and the mobile system.

4.1 The recommendation engine

The MNEMOSYNE prototype uses two different solutions to provide recommendations to the users: a knowledge-based and an experience-based system. These modules have been developed as web servlets which expose recommendation 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 ten 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 an RDF (Resource Description Framework) ontology that models all available multimedia content, see section 3.1 for details.

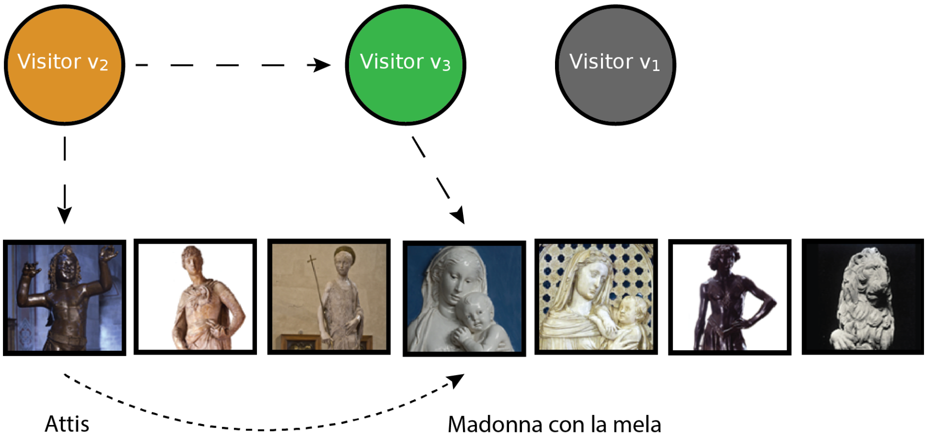


Fig. 5 Difference between user-based (large dashes) and item-based (short dashes) recommendation.

We use the SPARQL Protocol and RDF Query Language (SPARQL) 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 experience-based recommendation algorithms. The Recommender Engine implements metrics based on both user 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 described in section 3, 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 preference values and weights 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,[4] the state of the art for machine learning on big data.

A user-based recommender algorithm is based on the intuition that if two persons both prefer some of the same items, it is reasonable to assume that they will share many other preferences. Imagine the following simplified scenario: in the first 30 minutes since the museum opening there were three visitors (v_1, v_2 and v_3) in the Donatello Hall. During their visit a profile of interest has been passively built for each of them by the passive profiling module. For each visitor, we gather the top three artwork interests (ordered by the normalized value of interest noted $I_v(a)$ estimated as described in equation 6): for the first visitor we get $I_{v_1}(a_1) = 0.33$, $I_{v_1}(a_6) = 0.29$ and $I_{v_1}(a_5) = 0.23$; for the second visitor we get $I_{v_2}(a_3) = 0.33$, $I_{v_2}(a_9) = 0.32$ and $I_{v_2}(a_1) = 0.23$; and for the third visitor we

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

(a) (b)

Fig. 6 (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 recommendation space in the related resources level: information is related thematically to the selected artwork.

get $I_{v_3}(a_7) = 0.33$, $I_{v_3}(a_4) = 0.26$ and $I_{v_3}(a_1) = 0.26$, where a_1 corresponds to the artwork *Attis*, a_3 to *Madonna con bambino*, a_4 to *David* (Marble), a_5 to *Marzocco*, a_6 to *San Giorgio*, a_7 to *Madonna con la mela* and a_9 to *Sacrificio di Isacco*.

When visitor v_2 approaches the table, one of the recommendations proposed by the experience-based recommendation system is the *Madonna con la mela*. The reason can be deduced from the inner-workings of the recommendation algorithm. First of all, the module searches for the set of all artworks for which v_2 has no estimated preference. Then it searches all the users which have at least one preference on this set of artworks (both visitor v_1 and user v_3), computes a similarity between v_2 and them, and incorporates the preference for each artwork in the set, weighted by user similarity, into a running average. Considering artwork preferences, note that both user v_1 and user v_2 have *Attis* as a common preference but also that the preference value distance is shorter between user v_3 and v_2 (0.26 vs 0.23) than between user v_1 and v_2 (0.33 vs 0.23). Analyzing this data, the module has inferred that v_2 is more similar to user v_3 than to user v_1. So it was more likely that v_2 was recommended with one or more of user v_2's preferred artworks, like *Madonna con la mela*, once in front of the interactive table. Note that this is just an example and that in reality the recommendation system relies on many profiles of interest to statistically infer user-based recommendations.

The item-based recommendation module follows the same approach. The main difference is that similarity evaluation is computed between items instead of users. That is, the module compares series of preferences extracted for many users, for one item, rather than for many items by one user. Ultimately the user-based recommender finds similar users, and sees what they like, while the item-based one checks what the user likes, and then finds similar artworks. See figure 5 for an illustration of this recommendation example and of the difference between user-based and item-based recommendation.

Fig. 7 The information widget window: when the user selects an item from one space, multimedia content is displayed in the center of the screen. The artwork selected in the artwork level is visualized on the left with its title, thumbnail, location and level of visitor interest.

4.2 The tabletop system

The tabletop display is a large touchscreen device (a 55-inch display with Full HD resolution). When the passive profiling system detects a visitor approaching the table, it sends the interest profile to the user interface software, which then exchanges data with the recommendation system in order to load all the multimedia content that will be displayed for this user.

The metaphor proposed for the user interface is based on the idea of a hidden museum waiting to be unveiled, starting from the top (the physical artworks) and moving deeper towards additional resources such as explanations and relations between one artwork and others. The proposed metaphor aims at hiding the complexity of the data extracted by the recommendation and passive profiling systems by letting users make more limited and simpler actions in deciding which content to consume and to interact with.

Both vertical and horizontal navigation are used as a metaphor for strolling through the virtual multimedia art gallery. Users can move between the following levels of information:

The artwork level visualizes digital representations of the physical artworks for which the visitor has shown the highest level of interest based on the data created by the passive profiling system (see figure 6a). A vertical animation starts when the user touches an artwork item in order to move the point of view under the current space and reveal the next level including resources related to this artwork.

The related resources level is a horizontally-arranged space in which the visitor can navigate through the multimedia content related to the selected artwork. Related resources are organized into three categories:

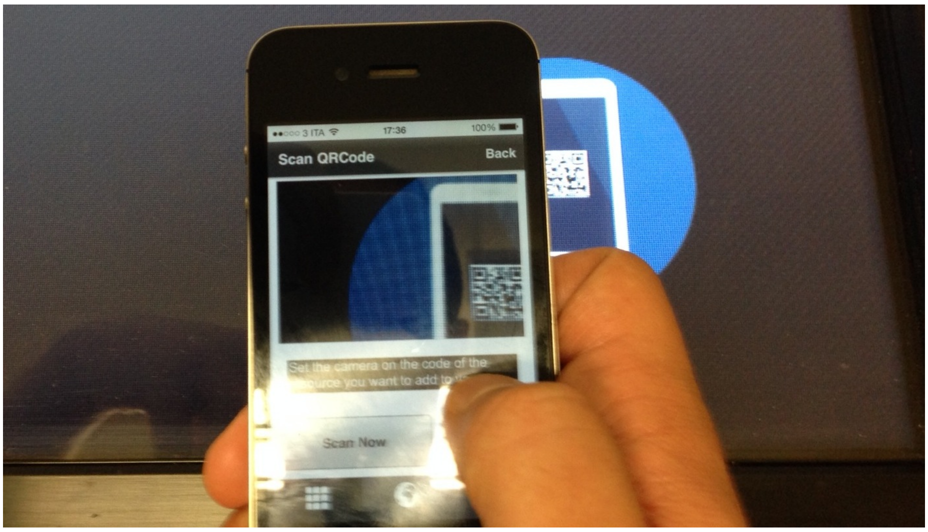


Fig. 8 The visitor can scan the QR code with the dedicated MNEMOSYNE mobile application in order to collect his personalized information about artworks.

– **insights:** stories directly related to the artwork in the ontology;
– **recommendations:** resources related to the artwork and its related stories in the ontology according to the knowledge-based recommendation system (see an example in figure 6b); and
– **social:** similar artworks according to the experience-based recommendation system using the visitor profile.

The interface makes the rules of navigation clear in these spaces by displaying a menu at the bottom of the screen which includes interactive textual labels. Each space is easily identified and memorized by users, because it is visualized with a different layout. The image used as background, the shape, size and positioning of items are all built to differentiate spaces of visualized data and to provide a pleasant user experience while navigating the resources related to an artwork. The *recommendations* space is shown in figure 6b. For each resource the user can activate a widget window which offers detailed information such as the title, the description and the author of the resource (see figure 7). According to this user interface and navigation paradigm, the visitor is able to easily find the most interesting information about artworks and related resources. Every interaction is recorded, and thus the user implicitly creates an active profile of interest in addition to the passive one obtained from the computer vision system. Both these profiles are available for later use via the mobile system detailed in the next section.

The software module that supports user interactivity with the tabletop was developed as a Rich Internet Application using Adobe AIR. The UI is based on Adobe Flash Builder and Action Script 3.0. QR code generation is implemented via the open source AS3 QR Code library.

4.3 Mobile system

As Ballagas et al. [6] point out, there are several issues concerning the interaction with displays in public spaces. They separate this kind of interaction into three application domains (personal, semi-public and public) and identify different constraints that should be considered for the design of these interaction techniques. The MNEMOSYNE project exploits both public (a large interactive tabletop) and private displays (visitor mobile devices) to provide information to visitors.

Furthermore, since the system provides personalized information, more considerations can be made when taking into account the utility of different devices. Rukzio et al. [45] underline that important advantages for users comes from storing personalized data on their mobile device. In fact, users can download information to their mobile phones which might be useful even after they have left the public display.

Taking these considerations into account, the main goal of the MNEMOSYNE mobile application, shown in figure 9, is to enable each visitor to collect personalized digital content displayed in the interactive tabletop interface. Therefore, unlike [17, 33], the mobile app is intended to be used at the end of the visit and not as an interactive device during the museum tour. It was developed using the Adobe AIR framework and can therefore be installed on devices running iOS (iPhone, iPad) or Android. In order to transfer data from the interactive display to the mobile device, we adopted a solution based on QR code scanning [28].

When the user touches the mobile icon at the bottom of the interface, the application running on the public interactive screen generates a QR code containing information about the user ID assigned by the computer vision system. Scanning the displayed code with the mobile application, as shown in figure 8, we transfer the unique user ID to the mobile device. The mobile application then queries the MNEMOSYNE database to retrieve the user's favorite artworks, generated both through the computer vision passive profiling module and from their interactions 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 allows us to extend the personalized user experience of the visit from an indoor to an outdoor scenario (see figure 9 right).

5 Field trials, evaluation and ongoing work

In this section we report on the ongoing validation of the MNEMOSYNE system in terms of profiling performance and usability.

5.1 Field trials

We are currently running field trials of the MNEMOSYNE prototype to evaluate its performance in in two different environments.

Permanent exhibition in Le Murate: We have installed a prototype of the MNEMOSYNE system in a permanent exhibition space at *Le Murate* in the his-

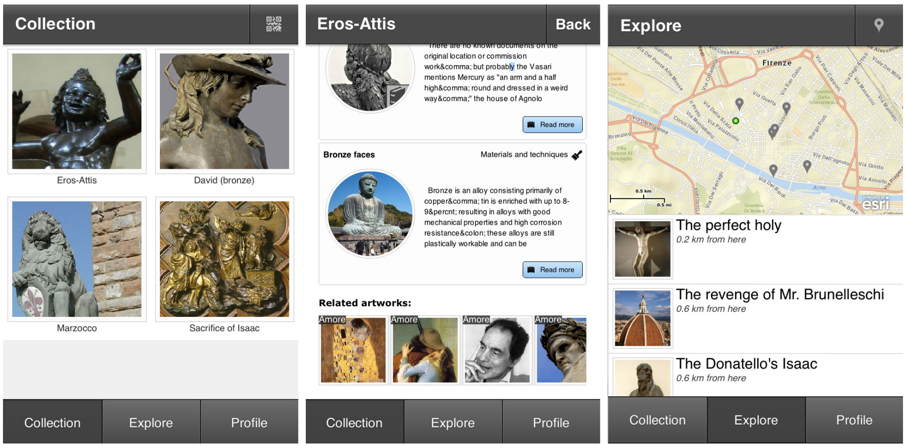


Fig. 9 The mobile application: the user's favorites artworks (on the left), in-depth information of an artwork (in the center), and the map of suggested points of interest (on the right).

torical 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. Examples of obtained detections are shown in figure 3a. 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.

Note that it would be possible to address the problem of lag due to the detector using a short-term tracking approach. Such an approach could estimate new positions and perform data association in the frames that would otherwise be skipped due to detector latency. However, such a tracking approach would require real-time, online multi-target tracking. We instead chose to directly address the problem by developing an adaptive detector that maximizes detection frame rate by automatically reducing the scales and positions to a small feasible set. We describe this detector briefly below.

Ongoing installation in the Bargello: The MNEMOSYNE project is currently being installed and beta-tested in the *National Museum of Bargello*. This installation uses four cameras, passively observing how visitors behave with respect to

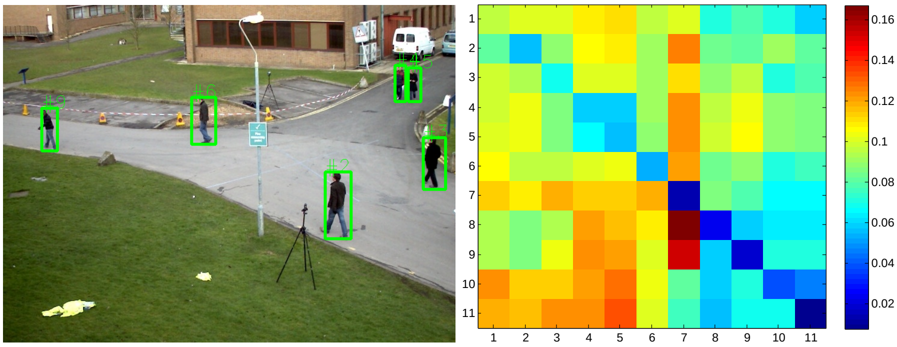


Fig. 10 Validation of the visual profiling system. (left) A frame with ground truth from the PETS dataset. (right) The distance matrix derived from our profiling system (see text for description).

the ten artworks selected in the *Sala Donatello*. The process of mapping the museum artworks and calibrating all cameras was done in half a day. The profiling system operates following the procedure describe in section 3: identity modeling is performed independently in each camera stream using Algorithm 1, and the final profiles of interest are obtained by merging local profiles.

The first experiments at *Le Murate* and preliminary results at the *Bargello* made clear that the main computational bottleneck of the profiling is due to the detection process. We hence developed a detector that learns only with weak supervision (the output of a rather slow pedestrian detector) where and at which scale detections usually appear [8]. At test time this detector will only generate candidate detection windows at scales and positions that are relevant for the image framed by each camera. In the next section we evaluate our profiling approach on a publicly available dataset.

5.2 Profiling validation

In the absence of annotated video from museum environments, we experimentally evaluated our proposed profiling approach on a dataset traditionally used for tracking and data association [4]. We ran an experiment on sequence S2.L1 of the PETS dataset. [5] The PETS dataset is appropriate for this type of evaluation because it includes typical, challenging situations including occlusions, people entering and exiting the scene, and people moving in groups. A frame from this sequence with ground truth identities is depicted in figure 10a. Our profiling approach relies on the set of detection positions associated with each identity. Hence, we evaluated our method by computing the Euclidean distance between the set of spatial positions of detections matching an identity and those given by the ground truth. More precisely, we compare for every frame the set of matched detections with the ground truth positions up to that frame. This measure is accumulated and then normalized by row. Eleven different identities are present in this sequence.

The results of this validation are given in the distance matrix in figure 10b. The overall lower concentration of energy on the diagonal of this matrix confirms

[5] `http://www.cvg.rdg.ac.uk/PETS2009/index.html`

that our identity modeling approach gives reasonable results. Note that this matrix even using the ground truth detections would not be perfectly diagonal. Indeed, two persons following similar path will have a small distance in this matrix. This is also how our approach behaves, as identities 4 and 5 for example evolve mostly side by side in the video the distances between their set of detections is small. This is actually a desireable behavior as it provides some robustness to "groups" of visitors who should have similar profiles of interest.

5.3 Usability study

The entire system is now being installed in its official location: the Hall of Donatello in the Bargello Museum in the city of Florence. To obtain preliminary, quantitative feedback on the beta version of the system, we organized a "testing week" during which we invited twenty-four visitors to use our system and then asked them to complete a questionnaire. We used a standard questionnaire based on the System Usability Scale (SUS) [7] in which all participants score the following ten items with one of five responses ranging from Strongly Agree to Strongly disagree:

1. I think that I would like to use this system frequently;
2. I found the system unnecessarily complex;
3. I thought the system was easy to use;
4. I think that I would need the support of a technical person to be able to use this;
5. I found the various functions in this system were well integrated;
6. I thought there was too much inconsistency in this system;
7. I would imagine that most people would learn to use this system very quickly;
8. I found the system very cumbersome to use;
9. I felt very confident using the system;
10. I needed to learn a lot of things before I could get going with this system.

The SU scale gives a global, subjective assessment of usability. The questionnaire is generally given after the user has used the system in an unsupervised way, before any debriefing or discussion takes place. Respondents are asked to record their immediate response to each item, rather than thinking about them for a long time. SUS scores have a range of 0 to 100. In addition, we added a few open-ended questions in order to collect suggestions from users.

Twenty-four visitors were willing to test the system and respond to our questionnaire (user ages ranged between 20 and 55 years, and 60% of users were male). Results of the SUS study are shown in figure 11. Note that in figure 11 the score given to each question is already reordered to give a higher value to a positive appreciation of the system.

The usability results for the beta version of the MNEMOSYNE system are encouraging. The average SUS score hovers around 73.5, indicating that this first functioning version of the system is already rather easily accepted by the users. The worst graded question was #5, with users finding various functions to be poorly integrated. This leads us to believe that some functions of the system may still be a bit confusing. However, the best results are obtained for question #8, with users apparently finding the interface intuitive. Moreover, the open-ended questions suggested some changes that will be included in future versions of the system.

SUS QUESTIONNAIRE

USER #	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	SUS SCORE
1	3	3	3	2	2	3	1	3	1	2	57.5
2	3	4	3	4	2	3	4	4	3	4	85.0
3	3	4	4	2	4	4	3	4	4	4	90.0
4	3	4	3	3	4	4	3	4	4	4	90.0
5	3	2	1	4	2	4	3	3	2	2	65.0
6	2	3	1	3	1	1	3	3	2	4	57.5
7	3	3	3	0	2	3	3	4	3	2	65.0
8	2	4	3	4	2	4	2	4	2	4	77.5
9	3	3	3	3	2	3	3	4	4	4	80.0
10	3	4	4	2	2	3	4	4	3	4	82.5
11	4	4	4	2	4	4	4	4	4	3	92.5
12	1	4	4	4	1	1	4	4	3	4	75.0
13	4	3	3	4	3	3	4	4	3	4	87.5
14	4	4	3	3	2	3	3	4	3	4	82.5
15	4	3	2	4	2	3	2	3	2	2	67.5
16	1	1	2	1	2	4	2	4	2	3	55.0
17	4	2	3	1	2	4	2	3	3	2	65.0
18	4	2	2	1	3	4	2	3	3	3	67.5
19	2	2	3	3	1	2	2	3	2	3	57.5
20	4	4	4	2	2	4	4	4	4	2	85.0
21	4	4	4	4	4	4	4	4	4	4	100.0
22	3	4	3	1	3	1	2	4	2	4	67.5
23	3	1	1	3	2	3	2	2	2	1	50.0
24	2	2	2	4	2	2	2	3	2	4	62.5
Average	3.00	3.08	2.83	2.67	2.33	3.08	2.83	3.58	2.79	3.21	73.54

SUS STATISTICS

SUS SCORE	RESULTS
AVERAGE	73.54
STANDARD DEVIATION	13.83
<= 60	5
60 - 70	7
70 - 80	2
80 - 90	6
>= 90	4

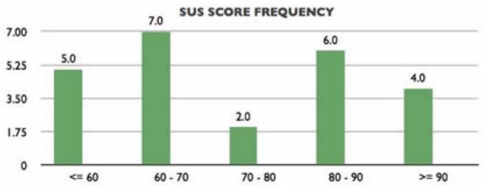


Fig. 11 Usability evaluation results based on the SUS questionnaire.

6 Conclusions

In this paper we detailed our proposal for a personalized multimedia museum experience. Our system makes use of passive observation of visitors in museum exhibits to independently estimate each visitor's interest profile. This profile is then used in conjunction with a recommendation system to provide personalized content delivery through a natural interaction interface on a large interactive table.

MNEMOSYNE is operational, has been tested in our exhibition space, and is currently under deployment inside the *National Museum of Bargello*. The first experiments and user studies demonstrate that the system does address each of its goals effectively. For the future, we are interested in evaluating how suggested resources impact visitor visits within one museum. Moreover, we are interested in using the mobile application as a bridge between different museums using the MNEMOSYNE system through suggestions of artworks of potential interest in other museums.

For future installations of the MNEMOSYNE system we plan to investigate techniques for automatically mapping the museum and artworks of interest using a moving RGB-D sensor. The automatic mapping of an image onto the RGB reconstruction could then be used to obtain groundplane positions of each artwork. The only supervision required in such a scenario would be labeling artworks in example images, plus some additional information about scale and zone of influence. This approach might also be usable for camera calibration, rendering setup and installation much simpler.

Acknowledgements

This work was partially supported by Thales Italia and the MNEMOSYNE project (POR-FSE 2007-2013, A.IV-OB.2). Andrew D. Bagdanov acknowledges the support of Ramon y Cajal Fellowship RYC-2012-11776.

References

1. Anderson, P., Blackwood, A.: Mobile and pda technologies and their future use in education. In: Proceedings of JISC Technology and Standards Watch, vol. 4, pp. 3–33 (2004)
2. Avraham, T., Gurvich, I., Lindenbaum, M., Markovitch, S.: Learning implicit transfer for person re-identification. In: Proceedings of ECCV - Workshops and Demonstrations, pp. 381–390 (2012)
3. Baber, C., Bristow, H., Cheng, S.L., Hedley, A., Kuriyama, Y., Lien, M., Pollard, J., Sorrell, P.: Augmenting museums and art galleries. In: 13rd Conference on Human-Computer Interaction Interact 2001, pp. 439–447 (2001)
4. Bagdanov, A.D., Del Bimbo, A., Di Fina, D., Karaman, S., Lisanti, G., Masi, I.: Multi-target data association using sparse reconstruction. In: Proc. of International Conference on Image Analysis and Processing (ICIAP), pp. 239–248 (2013)
5. 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)
6. Ballagas, R., Rohs, M., Sheridan, J.G., Borchers, J.: Byod: Bring your own device. In: UbiComp 2004 Workshop on Ubiquitous Display Environments (2004)
7. Bangor, A., Kortum, P.T., Miller, J.T.: An empirical evaluation of the system usability scale. Intl. Journal of Human–Computer Interaction **24**(6), 574–594 (2008)
8. Bartoli, F., Lisanti, G., Karaman, S., Bagdanov, A.D., Bimbo, A.D.: Unsupervised scene adaptation for faster multi-scale pedestrian detection. In: Proceedings of the 22nd International Conference on Pattern Recognition (ICPR 2014) (2014)
9. 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 (2006)
10. Bazzani, L., Cristani, M., Perina, A., Murino, V.: Multiple-shot person re-identification by chromatic and epitomic analyses. Pattern Recognition Letters **33**(7), 898–903 (2012)
11. Billsus, D., Pazzani, M.J.: User modeling for adaptive news access. User modeling and user-adapted interaction **10**(2-3), 147–180 (2000)
12. Bimbo, A.D., Lisanti, G., Masi, I., Pernici, F.: Person detection using temporal and geometric context with a pan tilt zoom camera. In: 20th International Conference on Pattern Recognition (ICPR), pp. 3886–3889. IEEE (2010)
13. Bowen, J.P.: A brief history of early museums online. The Rutherford Journal (2010)
14. Bowen, J.P., Filippini-Fantoni, S.: Personalization and the web from a museum perspective. In: Proceedings of the Museums and the Web Conference (MW2004) (2004)
15. Breitenstein, M.D., Reichlin, F., Leibe, B., Koller-Meier, E., Van Gool, L.: Robust tracking-by-detection using a detector confidence particle filter. In: 12th International Conference on Computer Vision, pp. 1515–1522 (2009)
16. Breitenstein, M.D., Reichlin, F., Leibe, B., Koller-Meier, E., Van Gool, L.: Online multi-person tracking-by-detection from a single, uncalibrated camera. IEEE Transactions on Pattern Analysis and Machine Intelligence **33**(9), 1820–1833 (2011)
17. Bruns, E., Brombach, B., Zeidler, T., Bimber, O.: Enabling mobile phones to support large-scale museum guidance. MultiMedia, IEEE **14**(2), 16–25 (2007)
18. Brusilovsky, P., Maybury, M.T.: From adaptive hypermedia to the adaptive web. Communications of the ACM **45**(5), 30–33 (2002)
19. Cheng, D.S., Cristani, M., Stoppa, M., Bazzani, L., Murino, V.: Custom pictorial structures for re-identification. In: Procedings of BMVC, vol. 2, p. 6 (2011)
20. 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 (2005)

21. Delgado, J., Ishii, N.: On-line learning of user preferences in recommender systems. In: Proc. of International Joint Conference on Artificial Intelligence (IJCAI-99), Workshop on Machine Learning for Information Filtering, Stockholm, Sweden (1999)
22. Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: Person re-identification by symmetry-driven accumulation of local features. Proceedings of CVPR pp. 2360–2367 (2010)
23. Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part-based models. Pattern Analysis and Machine Intelligence, IEEE Transactions on **32**(9), 1627–1645 (2010)
24. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision, second edn. Cambridge University Press, ISBN: 0521540518 (2004)
25. 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)
26. Huang, Z., Chen, H., Zeng, D.: Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering. ACM Transactions on Information Systems (TOIS) **22**(1), 116–142 (2004)
27. Hyvönen, E.: Semantic portals for cultural heritage. In: Handbook on ontologies, pp. 757–778. Springer (2009)
28. Jin, C., Takahashi, S., Tanaka, J.: Interaction between small size device and large screen in public space. In: Knowledge-Based Intelligent Information and Engineering Systems, pp. 197–204. Springer (2006)
29. 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)
30. Karaman, S., Bagdanov, A.D., D'Amico, G., Landucci, L., Ferracani, A., Pezzatini, D., Del Bimbo, A.: Passive profiling and natural interaction metaphors for personalized multimedia museum experiences. In: New Trends in Image Analysis and Processing–ICIAP 2013, pp. 247–256. Springer Berlin Heidelberg (2013)
31. Karaman, S., Lisanti, G., Bagdanov, A.D., Bimbo, A.D.: Leveraging local neighborhood topology for large scale person re-identification. Pattern Recognition (2014). DOI http://dx.doi.org/10.1016/j.patcog.2014.06.003. In press
32. 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, Advances in Computer Vision and Pattern Recognition, pp. 287–307. Springer (2013)
33. 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)
34. Kumar, R., Raghavan, P., Rajagopalan, S., Tomkins, A.: Recommendation systems: A probabilistic analysis. In: Foundations of Computer Science, 1998. Proceedings. 39th Annual Symposium on, pp. 664–673 (1998)
35. Leibe, B., Schindler, K., Cornelis, N., Van Gool, L.: Coupled object detection and tracking from static cameras and moving vehicles. Pattern Analysis and Machine Intelligence, IEEE Transactions on **30**(10), 1683–1698 (2008)
36. Leibe, B., Seemann, E., Schiele, B.: Pedestrian detection in crowded scenes. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, vol. 1, pp. 878–885 (2005)
37. Mikolajczyk, K., Schmid, C., Zisserman, A.: Human detection based on a probabilistic assembly of robust part detectors. In: Computer Vision-ECCV 2004, pp. 69–82 (2004)
38. Negroponte, N.: Being digital. Random House LLC (1996)
39. Norman, D.A.: The invisible computer: why good products can fail, the personal computer is so complex, and information appliances are the solution. MIT press (1999)
40. Papageorgiou, C., Poggio, T.: Trainable pedestrian detection. In: Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on, vol. 4, pp. 35–39 (1999)
41. Pechenizkiy, M., Calders, T.: A framework for guiding the museum tours personalization. In: Proceedings of the Workshop on Personalised Access to Cultural Heritage (PATCH07), pp. 11–28 (2007)
42. Popescul, A., Pennock, D.M., Lawrence, S.: Probabilistic models for unified collaborative and content-based recommendation in sparse-data environments. In: Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence, pp. 437–444 (2001)

43. Prosser, B., Zheng, W.S., Gong, S., Xiang, T.: Person re-identification by support vector ranking. In: Proceedings of BMVC (2010)
44. Rashid, A.M., Albert, I., Cosley, D., Lam, S.K., McNee, S.M., Konstan, J.A., Riedl, J.: Getting to know you: learning new user preferences in recommender systems. In: Proceedings of the 7th international conference on Intelligent user interfaces, pp. 127–134 (2002)
45. Rukzio, E., Schmidt, A., Hussmann, H.: An analysis of the usage of mobile phones for personalized interactions with ubiquitous public displays. In: Workshop Ubiquitous Display Environments in conjunction with UbiComp (2004)
46. Sakamura, K.: Personalized digital museum assistant. Digital Museum 2000 (2003)
47. Sampaio, I., Ramalho, G., Corruble, V., Prudêncio, R.: Acquiring the preferences of new users in recommender systems-the role of item controversy. In: Proceedings of the ECAI 2006 Workshop on Recommender Systems, pp. 107–110 (2006)
48. Savia, E., Puolamäki, K., Sinkkonen, J., Kaski, S.: Two-way latent grouping model for user preference prediction. In: In Proceedings of the UAI'05. Citeseer (2005)
49. Sparacino, F.: The museum wearable: real-time sensor-driven understanding of visitors' interests for personalized visually-augmented museum experiences. In: Proceedings of Museums and the Web (MW2002) (2002)
50. Sparacino, F.: Scenographies of the past and museums of the future: from the wunderkammer to body-driven interactive narrative spaces. In: Proceedings of the 12th annual ACM international conference on Multimedia, pp. 72–79. ACM (2004)
51. Starner, T., Mann, S., Rhodes, B., Levine, J., Healey, J., Kirsch, D., Picard, R.W., Pentland, A.: Augmented reality through wearable computing. Presence: Teleoperators and Virtual Environments **6**(4), 386–398 (1997)
52. Stock, O., Not, E., Zancanaro, M.: Intelligent interactive information presentation for cultural tourism. In: Multimodal Intelligent Information Presentation, pp. 95–111. Springer (2005)
53. Ungar, L.H., Foster, D.P.: Clustering methods for collaborative filtering. In: AAAI Workshop on Recommendation Systems, vol. 1 (1998)
54. Wang, X., Han, T.X., Yan, S.: An hog-lbp human detector with partial occlusion handling. In: IEEE 12th International Conference on Computer Vision, pp. 32–39 (2009)
55. Wang, Y., van Sambeek, R., Schuurmans, Y., Aroyo, L., Stash, N., Rutledge, L., Gorgels, P.: Be your own curator with the chip tour wizard. In: Proceedings of the Museums and the Web Conference (MW2008) (2008)
56. 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), 139–153 (2009)
57. Artabus. URL `http://www.artabus.com/`. Accessed: 2014-03-18
58. Art.net. URL `http://www.art.net/`. Accessed: 2014-03-18
59. Peach home page. URL `http://peach.fbk.eu/`. Accessed: 2014-04-07
60. Polo museale fiorentino :: Sito ufficiale. URL `http://www.uffizi.firenze.it`. Accessed: 2014-03-18
61. The san francisco museum of modern art (sfmoma). URL `http://www.sfmoma.org/`. Accessed: 2014-03-18
62. Site officiel du musée du louvre. URL `http://www.louvre.fr`. Accessed: 2014-03-18
63. Tate. URL `http://www.tate.org.uk/`. Accessed: 2014-03-18
64. Virtual museum exhibit home page. URL `http://www.virtualmuseumexhibit.com/`. Accessed: 2014-03-18
65. Web gallery of art. URL `http://www.wga.hu/`. Accessed: 2014-03-18
66. Wu, B., Nevatia, R.: Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors. In: 10th IEEE International Conference on Computer Vision, vol. 1, pp. 90–97 (2005)
67. Zancanaro, M., Stock, O., Alfaro, I.: Using cinematic techniques in a multimedia museum guide. In: Proceedings of Museums and the Web 2003 (2003)