A Technological Framework for Personalized Museum Visiting*

Massimiliano ALBANESE, Antonio PICARIELLO and Antonio Maria RINALDI
{malbanes,picus,amrinald}@unina.it
Dipartimento di Informatica e Sistemistica, University of Napoli
Napoli, Italy

ABSTRACT

Visiting museums and archeological sites is usually an amazing experience but sometimes a visitor may get lost in huge exhibitions in which thousands of artifacts are exposed. In this paper we present a system that assists a user in visiting a museum or an archaeological site, providing personalized recommendations, based on an innovative strategy for predicting user behavior. Our approach makes use of an artificial neural network trained on a suitable set of features characterizing the objects of interest and also takes into account site topological information gathered from a *Geographical Information System*.

Keywords: ANN, GIS, GPS, Location Aware Computing, Mobile/Wireless Applications.

1. INTRODUCTION

Visiting museums and archeological sites is usually an amazing experience but sometimes a visitor may get lost in huge exhibitions in which thousands of artifacts are exposed.
Paper or human guides may help to reach interesting items, but the offer cannot be personalized according to the preferences of each user. Paper guides are the same for each visitor while human guides usually drive a whole group of tourists. In order to make visitor experience in the museum or archaeological site more interesting and stimulating, the access to the exposed items should be differentiated according to the visitor's specific profile.
In this paper we present a system that assists a user in visiting a museum or an archaeological site, providing useful recommendations, based on an innovative strategy for predicting user behavior.

*This work has been carried out partially under the financial support of the Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) in the framework of the FIRB Project "Middleware for advanced services over large-scale, wired-wireless distributed systems (WEB-MINDS)"

At the best of our knowledge, no much work has been done in developing a technological framework to enhance visitor's experience by personalizing the access to collections of artifacts. In [14] a user-centered approach for computational storytelling is presented. The author describes the *Museum Wearable*, a device which delivers an audiovisual narration interactively in time and space to the visitor, depending on the estimated visitor type. The authors of [1] present a system that automatically clusters GPS data that are incorporated into a Markov model that can be consulted for use with a variety of applications in both single-user and collaborative scenarios. In particular the model is used for predicting user's movements.
Several approaches have been developed in order to simplify and personalize browsing and retrieval in large multimedia databases. Such techniques have been successfully applied in the scenario of virtual museums, i.e. museums that offer a web based access to a collection of digital reproductions of artifacts, mainly paintings. Drummond et al. [5] propose a technique to assist the users in their search through a multimedia database, based on the intelligent agents paradigm. In [3] Bodendorf et al. present a system architecture for hypermedia applications that includes fuzzy logic and artificial neural networks for dynamically creating user–specific paths through a database of multimedia objects.
The strategy we propose in our approach makes use of an artificial neural network, for predicting users' behavior, trained on a suitable set of features characterizing the objects of interest, and also takes into account site topological information gathered from a *Geographical Information System* (GIS).
In our application domain a great accuracy in the measure of position is a fundamental aspect because the objects of interest are usually very close each other. Standard GPS technology is affected by several classes of errors [12] deriving from transmitted location of the satellite (*Ephemeris data*), transmitted clock (*Satellite clock*), corrections of pseudorange caused by ionospheric and tropospheric effects, reflected signals entering the receiver antenna (*Multipath*) and errors in

the receiver's measurement of range caused by thermal noise, software accuracy, and inter-channel biases. These errors can be eliminated or reduced using the differential correction technique (DGPS), which employs a second receiver at a fixed location to compute the corrections of the GPS satellite measurements. In this way we can improve the accuracy of the measure from $15-20$ meters up to less than $2-3$ meters. The fixed receiver has a radio connection (beacon) to transmit corrections to the devices, or alternatively, uses an internet connection.

In outdoor positioning, GPS [11] is used in many applications. A restriction for GPS technology is the impossibility to evaluate position in indoor spaces, where GPS receivers are blind. We can use several kinds of sensors such as infrared (IR) [16], ultrasound [13, 17], vision [9] or radio (RF) sensors [2, 7, 10, 18] to determinate the position in a indoor environment. These systems differ for many parameters as sensors used, cost of system components, hardware, time and space resolution [6]. Many issues arise in developing a robust location system in an indoor environment. These restrictions consist for example of an exact knowledge of spatial position of sensors and errors in the propagation of wave signals [4].

The exponential growth in the use of wireless communication drives the developers of mobile devices to equip their products with off-the-shelf IEEE 802.11b wireless Ethernet; at the same time the communication infrastructures of many buildings have often base stations based on the same IEEE 802.11*b* protocol. The localization is obtained by the measure of *RF* signal strength.

The model behind our system might be easily extended to any kind of museum, even if in this paper we make some assumptions that are valid in the case of a museum exhibiting paintings.

The rest of the paper is organized as follows. Section 2 shows the overall architecture of the proposed system. Sections 3, 4, and 5 describes the GIS *Support*, the *Prediction* and the *Path Planning* subsystems respectively. In section 6 some conclusion are eventually reported.

2. SYSTEM ARCHITECTURE

Figure 1 shows at a glance the overall architecture of the system. A user who wants to visit the site only needs a hand held computer, equipped with a positioning system and a wireless Ethernet connection, to access the *Recommendation System* and start its personalized visit. The positioning system allows to track the position of the visitor, that is periodically transmitted to the server, and stored in the *Usage Log*. The *Prediction Subsystem*, based on the history of the user and the behavior of the past users, predicts which are the items that it may be interested to see next. The *Path Planning Subsystem* proposes a path to the user, taking into account the results of the *Prediction Subsystem* and the information provided by the GIS *Support Subsystem*.

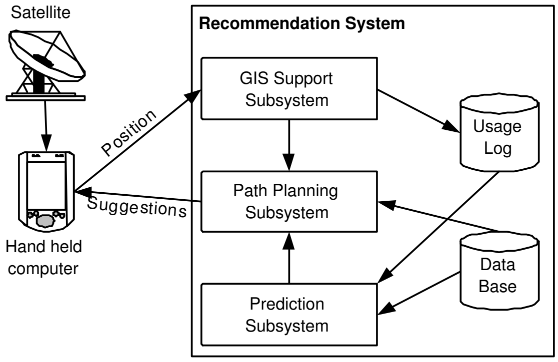

Figure 1: System architecture

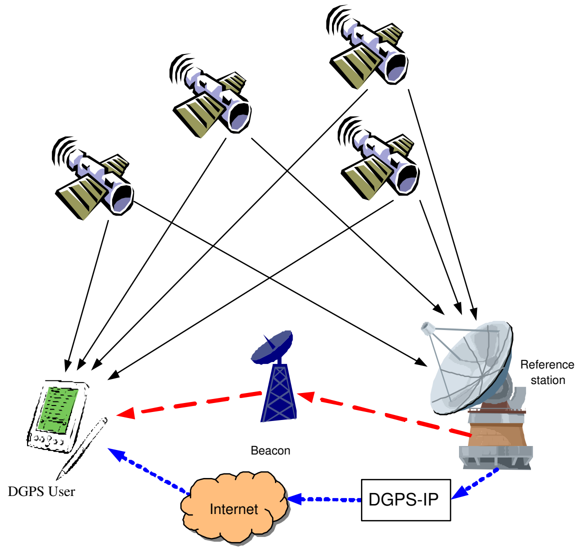

Figure 2: DGPS

3. GIS SUPPORT SUBSYSTEM

The data analysis on large repository storing several kind of data allows us to combine, transform and retrieve relations among the data themselves in order to establish a reliable decision support methodology. In our framework the use of spatial and temporal information is a strategic aspect for real-time applications. In particular, GIS supplies a useful technological support to manage this kinds of data. On the other hand DGPS technology can be used for tracking the user position in a reliable and accurate way, as shown in figure 2.

We use IEEE 802.11b wireless Ethernet for localization and usual communication services (internet ac-

cess). The use of more sophisticated localization systems, as example sonar, that allows to achieve a better performance (up to 20 centimeters), may be subject to environmental and legal constraints for cultural heritage buildings. In our system we get information about the user position and the time when that position is measured, and then interpolate the collected data in order to reconstruct the user's path and compute the spatial proximity to the objects in the site.
Our GIS module implements some functions to manage the geographic path of a user. We define a net where nodes are the objects and arcs are the connection among them; the weight on each arc is the distance among objects and around each node we consider a buffer area to define the *object's space of interest*: if a visitor is in this area we assume that the object is of interest to it. The distance between two objects o_i and o_j is defined as

$$dist_s(o_i, o_j) = \min_h \{\omega_h(o_i, o_j)\} \tag{1}$$

where ω_h are the paths between o_i and o_j. We can evaluate the time spent by a user looking an artifact by considering the points in the object's buffer. We select the points in the buffer area using the GIS module with a spatial join between buffer and points. This operator allows to extract user information about proximity to the artifacts and visit time that can be evaluated as the difference between the time of first and the last position points along its path that fall in the buffer.

4. PREDICTION SUBSYSTEM

In this work we are interested in predicting the objects that a visitor of a museum may be interested to visit, based on its behavior and on the behavior of the past visitors. Such prediction can be used for both making useful suggestions to the visitors and pre–fetching the description of the predicted objects, thus improving the performance of the system.
No mechanisms such as cookies or explicit user login have been implemented to simplify the task of user identification and classification, since the first ones can be deleted or disabled by the user itself and the same device might be used by several users over time, while the explicit login and the typing of personal data typically discourages the visitors from using the service, even if it is regarded as interesting. So the precision of user classification, being exclusively based on his dynamic behavior, is quite poor when the user accesses the system for the first time and then it gets better and better as it keeps on visiting the museum.
Knowledge about the visitors' behavior can be derived by the analysis of the *Usage Log* that records which artifacts have been visited by the users and how much time they have spent in front of each item. Assuming that $\mathcal{O} = \{o_i\}$ is the set of all the artifacts exposed in the museum or archeological site, let us introduce the following definition.

Definition 1 (Timestamped Usage Path) *A* Timestamped Usage Path *(TUP) p of length k is an ordered sequence of $(o_i, t_i) \in \mathcal{O} \times \mathbb{R}^+$ pairs, where o_i is the i-th object visited by the user in the same tour and t_i is the time that it has spent in front of o_i.*

$$p = ((o_{i_1}, t_{i_1}), ..., (o_{i_k}, t_{i_k})) \tag{2}$$

Let $\mathcal{P}$ denote the set of all the timestamped usage paths of past visitors stored in the log. We remark that the times are evaluated as described in section 3. To the aim of prediction we have adopted an *Artificial Neural Network* (ANN) approach, designing a neural network that is trained on a suitable set of features characterizing the objects of interest. We describe the features extraction process and the design of the neural network in the remainder of this section.

Features extraction
We assume that each object o in the collection $\mathcal{O}$ has a digital representation $I = r(o)$, that is usually a picture of the object itself. A feature vector $v_f^I = (f_1, ..., f_n)$ in an n-dimensional space can be extracted from each object representation and a distance $dist_f$ can be defined in the feature space. In the following we report two examples of feature extraction w.r.t. the image domain; suitable distances are respectively defined too. We have adopted the Wavelet Transform (WT) [15] as a mechanism useful for both reducing the amount of data to be analyzed and providing a suitable color and texture representation.

Color features: Given a set of representative colors $Q = \{q_1, ..., q_B\}$, a color histogram $h(I) = \{h_b^I\}$ of an image I is defined on bins $b \in [1, B]$, such that, for any pixel in D_p, h_b^I is the probability that the color of the pixel is $q_b \in Q$. To this aim, we have used the low pass component of the WT, i.e. a smoothed copy of the original picture, that allows to avoid lightening and noise problems.

Definition 2 (Color Distance) *Given two images I_1 and I_2 and their respective color histograms $h(I_1) = \{h_b^{I_1}\}$ and $h(I_2) = \{h_b^{I_2}\}$, defined on the same number B of bins, a* Color Distance *can be defined as*

$$dist_c(I_1, I_2) = 1 - \sum_{b=1}^{B} \min\left(h_b^{I_1}, h_b^{I_2}\right) / \sum_{b=1}^{B} h_b^{I_1}, \tag{3}$$

where $\sum_{b=1}^{B} h_b^{I_1}$ is a normalization factor.

Texture features: Let us denote the wavelet coefficients as $w_l^k(x, y)$, where $(x, y) \in D_p \subseteq \mathbb{R}^2$, l is the decomposition level and k the sub-bands. A wavelet decomposition gives rise to 4 subregions of dimension $|D_p|/4$. Only the detail components of the WT are taken into account, in order to characterize texture. For $k = 1, 2, 3$, the detail sub-bands contain horizontal, vertical and diagonal directional information, respectively, and are represented by coefficient planes $\left[\{w_l^k(x,y)\}\right]_{k=1,2,3}$. Next, the Wavelet Covariance Signature is computed, i.e. the feature vector of coefficient covariances $\Sigma_C^2 = \{\sigma_{X,Y}^2\}$, where:

$$\sigma_{X,Y}^2 = \sum_{x,y} \left\{ \frac{1}{|D_p|/4} \sum_{k=1}^{3} X_k(x,y) Y_k(x,y) \right\}. \quad (4)$$

The pair (X_k, Y_k) is in the set of coefficient plane pairs $\{(w_i^k, w_j^k)\}$, i and j being used to index the three channels, and (x, y) span over the sub-band lattice of dimension $|D_p|/4$.

DEFINITION 3 (Texture Distance) *Let C_1 and C_2 be the wavelet signatures of two images I_1 and I_2 respectively. A* Texture Distance *between two images I_1 and I_2 can be defined as*

$$dist_t(I_1, I_2) = \frac{1}{R} \sum_{i=1}^{|\Sigma^2|} \frac{\left|\Sigma_{C_1}^2[i] - \Sigma_{C_2}^2[i]\right|}{\min\left(\left|\Sigma_{C_1}^2[i]\right|, \left|\Sigma_{C_2}^2[i]\right|\right)} \quad (5)$$

where R is a normalization factor to bound the sum in $[0, 1]$, and $|\Sigma^2|$ the number of features in the feature vector Σ^2 computed through equation 4.

In this work we have used a feature vector that contains both color and texture features, since they have been proved to be powerful descriptors for pictorial images. Thus we have introduced a distance that combines color and texture distances, as reported in the following.

DEFINITION 4 (Feature Distance) *The* Feature Distance *between two images I_1 and I_2 is defined as*

$$dist_f(I_1, I_2) = \alpha_c \cdot dist_c(I_1, I_2) + \alpha_t \cdot dist_t(I_1, I_2) \quad (6)$$

α_c and α_t being two weighting factors.

In the following of the paper, we will be using the term *object* to refer both to the artifacts in the museum and their digital representations; it will be clear from the context which meaning is intended.

Neural network implementation
Among the several ANN architectures presented in the literature we have chosen a feed–forward topology, that is well–suited for classification problems [8]. In feed-forward networks neurons are typically organized into layers. A standard L–layer feed-forward network[1] consists of an input stage, L–1 hidden layers and an output layers of unit successively connected in a feed-forward fashion with no connections between units in the same layer and no feedback connections.
In this work we have designed a 4–layer feed–forward network, as sketched in figure 3. Each unit in the first hidden layer defines an hyperplane in the pattern space. A unit in the second hidden layer defines a hyper–region from the outputs of the first-layer unit; a decision region is obtained by performing an AND operation on the hyperplanes. The units in the third layer combine the decision regions defined by the units in the second hidden layer by performing logical OR operations, thus allowing to define arbitrarily complex decision boundaries and represent any boolean function.
The inputs to the network are the feature vectors $v_f^i = (f_1^i, .., f_n^i)$ of the last N objects visited by the user and their respective visit times, while the output is a vector $v_f^{N+1} = (f_1^{N+1}, .., f_n^{N+1})$ containing the desired features of the objects that the user is likely to visit next.
Neurons in the three hidden layers have a sigmoid activation function $f(x) = 1/(1 + e^{-\alpha \cdot x})$, where α is a parameter that controls the slope of the curve. The sigmoid activation function allows to divide the $N \cdot (n + 1)$–dimensional input space into smooth decision regions rather than piecewise linear regions. Units in the output layer have a linear activation function that permits to obtain an output vector in the features space $\mathcal{F}$.
A standard back–propagation algorithm [8] has been used to learn the connection weights from available training patters, in a supervised fashion. The data set to be used in the learning process has been built by running the system without the support of the neural network and collecting the usage patterns of visitors over a sufficient time interval. All the subsequences of length $N + 1$ of the paths in $\mathcal{P}$ have been then considered, using the feature vectors of the first N element as the inputs and the feature vector of the last one as output.
Cross validation methodology has been used in order to choose the optimal number of neurons in the hidden layers and evaluate the generalization capability of the designed neural network.

5. PATH PLANNING SUBSYSTEM

We can finally define how to make suggestions to the visitors and plan their future path through the museum combining the information gathered from the GIS Support Subsystem and the results of the Prediction Subsystem. The basic consideration is that,

[1]We adopt the convention that the input nodes are not counted as a layer.

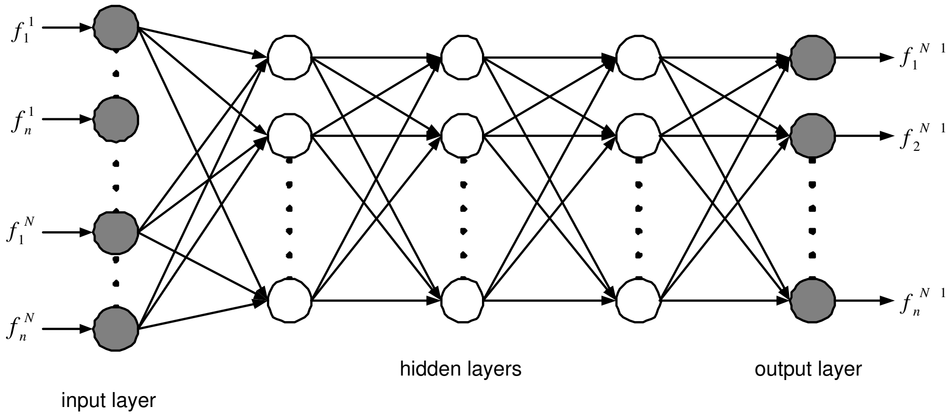

Figure 3: Neural Network for users' behavior prediction

when visiting a museum or an archaeological site, a user would like to see interesting artifacts, but, at the same time, it doesn't want to go up and down through the site to accomplish its cultural needs.
Given the last N objects $((o_{i_1}, t_{i_1}), ..., (o_{i_k}, t_{i_k}))$ visited by the user, the neural network returns a vector $v_f^{N+1} = (f_1^{N+1}, .., f_n^{N+1})$ containing the desired features of the objects that the user is likely to visit next. We can now define a distance d that combines the feature and spatial distances defined by Eq. 6 and Eq. 1 respectively.

$$d(o_j) = \alpha \cdot dist_f(o_d, o_j) + \beta \cdot dist_s(o_{i_N}, o_j),\ o_j \in \mathcal{O} \quad (7)$$

where o_d is an hypothetic object characterized by the desired features v_f^{N+1}, and α and β are two parameters used for both weighting the contribution of feature and spatial distances and making the two metric spaces coherent.
In conclusion, the path proposed to the user is a list of the top–k artifacts $\{o_j\}$ ranked by ascending values of $d(o_j)$.

6. CONCLUSIONS

In this paper we have presented a novel technological framework for personalizing museum visits and make visitors' experience more appealing. Our approach is based on an innovative strategy for predicting user behavior, that uses an artificial neural network trained on a suitable set of features characterizing the objects of interest and also takes into account site topological information (GIS).
So far the system has been implemented for outdoor environments, using the DGPS technology. We are planning to implement the indoor solution described in the paper and manage the integration of the two technologies, allowing the handover between the two tracking solutions, so that a user can walk within an outdoor environment and, at any time, enter a building without losing the connection with the system.

7. REFERENCES

[1] D. Ashbrook and T. Starner. Using gps to learn significant locations and predict movement across multiple users. *Personal Ubiquitous Comput.*, 7(5):275–286, 2003.

[2] P. Bahl and V. N. Padmanabhan. Radar: An in-building rf-based user location and tracking system. In *Proceedings of IEEE Infocom 2000*, volume 2, pages 775–784, Tel Aviv, Israel, March 2000.

[3] F. Bodendorf and K. Langer. Hypermedia navigation support by fuzzy logic and neural networks. In *Proceedings of IEEE International Conference on Intelligent Processing Systems*, pages 180–184, Beijng, China, October 1997.

[4] A. Chakraborty. *A distributed architecture for mobile, location-dependent applications.* PhD thesis, Massachusetts Institute of Technology, May 2000.

[5] C. Drummond, D. Ionescu, and R.Holte. Intelligent browsing for multimedia application. In *Proceedings of MULTIMEDIA'96*, pages 386–389, 1996.

[6] J. Hightower and G. Borriello. Location systems for ubiquitous computing. *IEEE Computer*, 34(8):57–66, August 2001.

[7] J. Hightower, R. Want, and G. Borriello. Spoton: An indoor 3d location sensing technology based on rf signal strength. Technical Report UW CSE 00-02-02, University of Washington, Department of Computer Science and Engineering, Seattle, WA, February 2000.

[8] A. Jain, J. Mao, and K. Mohiuddin. Artificial neural networks: A tutorial. *IEEE Computer*, 29(3):31–44, March 1996.

[9] J. Krumm, S. Harris, B. Meyers, B. Brumitt, M. Hale, and S. Shafer. Multi-camera multi-person tracking for easyliving. In *Proceedings of 3rd IEEE International Workshop on Visual Surveilance*, Dublin, July 2000.

[10] A. M. Ladd, K. E. Bekris, A. Rudys, G. Marceau, L. E. Kavraki, and D. S. Wallach. Robotics-based location sensing using wireless Ethernet. In *Proceedings of 8th ACM International Conference on Mobile Computing and Networking (MOBICOM)*, Atlanta, GA (USA), September 2002.

[11] T. Logsdon. *Understanding the Navstar: GPS, GIS and IVHS*. Van Nostrand Reinhold, New York, second edition, August 1995.

[12] B. W. Parkinson. *Global Positioning System: Theory and Applications*. American Institute of Aeronautics and Astronautics, June 1996.

[13] N. B. Priyantha, A. K. L. Miu, H. Balakrishnan, and S. J. Teller. The cricket compass for context-aware mobile applications. In *Proceedings 7th Annual ACM/IEEE International Conference on Mobile Computing and Networking (MOBICOM 2000)*, pages 1–14, Rome, Italy, July 2001.

[14] F. Sparacino. Sto(ry)chastics: a bayesian network architecture for user modeling and computational storytelling for interactive spaces. In *Proceedings of Ubicomp, The Fifth International Conference on Ubiquitous Computing*, Seattle, October 2003.

[15] G. Van de Wouwer, P. Scheunders, S. Livens, and D. Van Dyck. Wavelet correlation signatures for color texture characterization. *Pattern Recognition*, 32(3):443–451, March 1999.

[16] R. Want, A. Hopper, V. Falcão, and J. Gibbons. The active badge location system. *ACM Transactions on Information Systems*, 10:91–102, January 1992.

[17] A. Ward, A. Jones, and A. Hopper. A new location technique for the active office. *IEEE Personal Communications*, 4(5):42–47, October 1997.

[18] J. Werb and C. Lanzl. Designing a positioning system for finding things and people indoors. *IEEE Spectrum*, 35(9):71–78, September 1998.