senseable city lab:.::

New tools for studying visitor behaviours in museums: a case study at the Louvre

Yuji Yoshimura[a], Fabien Girardin[b], Juan Pablo Carrascal[a], Carlo Ratti[c] and Josep Blat[a]

[a] Information and Communication Technologies Department
Universitat Pompeu Fabra, Spain
[yuji.yoshimura, jp.carrascal, josep.blat] @upf.edu

[b] Lift Lab
Switzerland
fabien@liftlab.com

[c] SENSEable City Laboratory
Massachusetts Institute of Technology, USA
ratti@mit.edu

Abstract

In this paper we discuss the exploitation of data originated from Bluetooth-enabled devices to understand visitor's behaviour in the Louvre museum in Paris, France. The collected samples are analysed to examine frequent patterns in visitor's behaviours, their trajectory, length of stay and some relationships, offering new details on behaviour than previously available. Our work reinforces the emergence of a new methodology to study visitors. It is part of recent lines of investigation that exploit the presence of pervasive data networks to complement more traditional methods in tourism studies, such as surveys based on observation or interviews. However, most past experiments have explored quantitative data coming from mobile phones, GPS, or even geotagged user generated content to understand behaviour in a region, or a city, at a larger scale than that of our current work.

Keywords: Bluetooth sensing; human behaviour; museum study; real time management tool.

1 Introduction

In recent decades, tourism has developed to become one of the biggest industries. The World Tourism Organization foresees that the number of tourists will reach 1,600 million around 2020 and the World Travel and Tourism Council predicts that direct/indirect economic impact generated by the touristic industry will amount to 9.6% of Gross Domestic Product (GDP) and generate 9.7% of employment all over the world in 2012[1].

The increase of its economical, cultural and social impact on urban areas requires more precise and dynamic understanding of tourist behaviours and movements at micro (e.g. district, city) and macro (e.g. region, country) scales. Some emerging technologies make it possible to record and analyse them at city and district level (e.g. GPS, mobile phones with or without GPS (Asakura & Iryob, 2007); the passive mobile positioning data (Ratti, Pulselli, Williams, & Frenchman, 2006, Ahas, Aasa, Roose, Mark, & Silm, 2008); user-generated data (Girardin, Dal Fiore, Ratti, & Blat,

[1] http://www.wttc.org/eng/Tourism_Research/Economic_Research/

2008, Pereira, Vaccari, Girardin, Chiu, & Ratti, 2012, Girardin, Calabrese, Dal Fiore, Ratti, & Blat, 2008). In museums, the observation, and interview-based surveys have been used mostly to understand the social use of the environment and evaluate its use (see Hooper-Greenhill, 2006 for a review and Yalowitz & Bronnenkant, 2009, Hillier & Tzortzi, 2006). The information collected by these "traditional" methods provides support for the management of the spaces, which have a proved value. However they often provide a snapshot on the life of a built environment, and the interviews and questionnaires can have self-reporting bias. Moreover, they fail to record empirical evidences and measures (e.g. visiting time, sequences of visits, time of stay, density) key to produce a more complete picture on people use of a space.

The purpose of this paper is to discuss a Bluetooth proximity detection approach (previously developed for a traffic data collection system (Sanfeliu, Llácer, Gramunt, Punsola, & Yoshimura, 2010) to gather insights on visiting behaviours in a museum context and to demonstrate its relevance to support the management of environments that must respond to the increasing tourism demand. For instance, the analysis reveals the dynamic description of different use of museum spaces, the visiting profiles and the spatio-temporal patterns of visitors' behaviours.

In section 2 a brief summary of related works, their contributions and their main limitations is provided. In section 3 a Bluetooth proximity approach to detect visitor's presence and sequential movements is proposed. In section 4 the dataset processing for the analysis is discussed and key concepts for our research introduced. Section 5 presents some initial findings from our field trials, the frequent pattern and visitors' spatial uses. Finally, we summarize our on-going work on developing methods and tools for analysing the museum and urban environments.

2 Strategies to collect empirical visitor data

With the emergence of location technologies, a variety of methodologies have been proposed to locate a person, specifically for the collection of empirical data in the context of tourism (Asakura & Iryob, 2007, Ratti et al., 2006, Ahas et al., 2008, Girardin, Calabrese et al., 2008, Girardin, Dal Fiore et al., 2008, Yalowitz & Bronnenkant, 2009, Hillier & Tzortzi, 2006, Kanda et al., 2007). They are classified into 3 groups, and remark the burden which each method imposes on the persons involved.

The first group of more traditional techniques includes observation, and interviews. With the latter or with user diaries, one can obtain specimens of detailed visitor's behaviour. However, the data can be subjectively biased and the methods are costly requiring a lot of human resources (Girardin, Dillenbourg, & Nova, 2009). Something similar happens with direct observation, which could be difficult to sustain for long periods as it poses a heavy burden to the observer. The representativeness of the sample in interviews and questionnaires can be an issue too.

The second group is based on technologies such as GPS or RFID, which can supply more objective and precise time, location and route data (Asakura & Iryob, 2007, Kanda et al., 2007) than the traditional methods – however, without the motivations which can appear in interviews. Currently, these techniques demand the users to carry specifically enhanced devices that are not widespread. They make the data collection more cumbersome, and they may bias the user behaviour, and thus the collected data.

A third group includes using image sensing devices or passive mobile positioning data, which give little burden to the users – but no motivations are available either. Their main limitation is spatial; for instance, image sensing devices can record visitor's behaviour with spatio-temporal accuracy (Antonini, Bierlaire, & Weber, 2006), but the recording area covered by a single camera is limited (Yalowitz & Bronnenkant, 2009). Passive mobile positioning data have started to be used in tourism studies (Ratti et al., 2006, Ahas et al., 2008) as it can provide better empirical data on movement of people at global scale; nevertheless, the estimation of the presence and movement of people is limited by the cell size (i.e. the area of coverage of the base station that serves the mobile service).

Table 1. Data capture techniques showing their main strengths and weaknesses in the context of tourism and urbanism studies

Data capture	**Strengths**	**Weaknesses**	**Application example**
Manual surveys	Capture motivations	Very costly and applied to a limited time period	Timing and Tracking Survey (Yalowitz & Bronnenkant, 2009)
GPS and Cell phone (device-based)	Timely mobility data (potentially augmented with in-situ survey)	Survey limited in time and participants. It does not work inside the buildings	Describe social and spatial characteristics with limited samples (Asakura & Iryob, 2007)
RFID	Precise real-time mobility data	Survey limited in time and participants. Infrastructure deployment needed	Describe social and spatial characteristics with limited samples (Kanda et al., 2007)
Cell phone (aggregated network-based)	Use existing infrastructure to provide real-time mobility data	Does not work at the building and room scale	Real-time urban dynamics (Ratti et al., 2006)
Bluetooth detection	Precise real-time mobility data, non-intrusive to participants	Infrastructure deployment needed	Describe social and spatial characteristics (Kostakos, O'Neill, Penn, Roussos, & Papadongonas, 2010)

This paper presents several contributions in the development of data collection tools and methodologies for the analysis of large samples describing visitor's behaviour at small spatial scale using Bluetooth. The recent wide spread of mobile devices implies that many people have their Bluetooth switched on passively, thus providing an important source of useful data. A variety of projects have exploited Bluetooth data for measuring the social network relationships between people (Eagle & Pentland, 2005, Paulos & Goodman, 2004, Nicolai, Yoneki, Behrens, & Kenn, 2006), mobility of vehicles (Yalowitz & Bronnenkant, 2009, Barceló, Montero, Marqués, & Carmona, 2010) and mobility of pedestrians and their relationships (O'Neill et al., 2006, Kostakos et al., 2010). However these investigations have not considered a specific analysis of pedestrians and their use of space. This paper aims at reducing this shortcoming.

3 Data collection settings

A large majority of mobile devices currently on the market embed Bluetooth, and a significant proportion of users have them turned on in passive mode (Kostakos et al., 2010). The presence of these Bluetooth-enabled devices can be detected by means of sensors that scan the wireless spectrum. This section and the following describe the settings of our study and how the collected data were structured to handle privacy issues and allow for the spatial behavioural analysis.

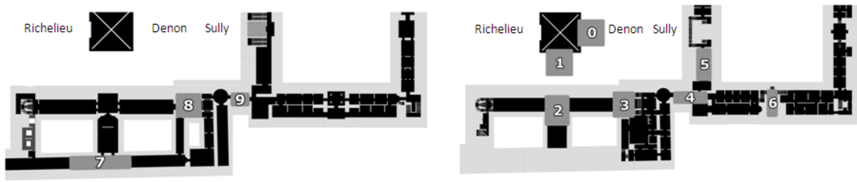
3.1 Context of the study

The Louvre is the most visited museum in the world with 8.5 million visitors in 2009 and more than 40,000 visitors at peak days[2]. This context of "cultural enthusiasm" has direct consequences on the quality of the visitor experience as well as on the organization and management of the Museum (e.g. application of flow management strategies and increased stress level of the surveillance staff). In response to the increasing tourism demand and the necessity to setup and evaluate museum strategies, we proposed to collect and analyse empirical data on the flows and occupancy levels of visitors in key areas of the Louvre.

3.2 Study settings and characteristics of the Bluetooth sensors

Because of the context, our study is particularly focused on one of the busiest areas of the museum identified by Le Louvre officials, namely a trajectory that leads visitors from the entrance (Pyramid) to the Venus de Milo. 10 Bluetooth sensors were deployed and they were sufficient to gather measures of visiting sequences and staying times at representative locations along the path (Figure 1). Two were on floor -1 (0 or *Hall*, 1 or *Denon access*); five were on floor 0 (2 or *Denon* 0, 3 or *Samothrace* 0, 4 or *Venus de Milo*, o 5 or *Caryatides*, 6 or *Sphinx*), and 3 on floor 1 (7 or *Big Gallery*, 8 or *Samothrace* 1, 9 or *Glass*).

The sensors gathered a unique encrypted identifier distinguishing each mobile device that supports Bluetooth and is set to be discoverable, as well as 2 time stamps for check-in and check-out times within the range of each sensor. Assuming that a mobile device belongs to a person, the movement of the device can be related to that of the visitor.



---
[2] http://www.theartnewspaper.com/attfig/attfig10.pdf

Fig. 1. Location of 10 sensors (No.0-No.9) indicating their approximate sensing range

The administrative and technical restrictions (e.g. protection against robbery, areas unreachable to visitors, no sources of electrical power, safety and health concerns) guided the deployment of the devices, sometimes preventing the installation in ideal locations for optimal detection. These special circumstances required the use of an ad-hoc battery with 10 days of autonomy for each sensor. This temporal limitation constrained the accumulation of empirical data, but our analyses show that this period is sufficient to extract relevant evidences.

3.3 Data and privacy issues

Based on a previous research that focused on the privacy issues related to the use of Bluetooth scanners (Sanfeliu et al., 2010), we adopted a solution that 1) does not allow the identification of individuals, 2) keeps the anonymity of trajectory data even after recording and archiving. This is achieved with the application of Secure Hash Algorithm (SHA) to the Bluetooth unique IDs detected by our system.

3.4 Sensor detectable area and the definition of its node

The spatial definition of the detectable area by a Bluetooth sensor is a critical issue for any research which uses this type of sensors. The shape of the area is similar to a flower with four petals of different length and width. In an optimal setting, the largest petal is an ellipse of almost 40 meters long by 15 meters wide, while the smallest is approximately 15 by 10 meters. The other two have a similar shape and a size of 15 by 10 meters. However, it could be customized for an indoors space with the largest petal dimensions being 20 by 7-8 meters. We identify the area detectable by a sensor as a node which represents the corresponding location, and use this definition through the rest of the paper. The detectable area estimations fluctuate according to the museum settings, due to the location of the sensors (e.g. within wooden boxes or administrative desks) and other factors, but we made sure that they would cover the targeted areas along the studied visitor trail.

Fig. 2. Conceptual diagram of Bluetooth sensor's detectable area

4 Collected data and measures

Based on the methodologies proposed in previous sections, the sample data during a specific audit period (more on this in section 4.2) was collected. This section describe how these amounts of collected sample data were organized to extract the desired values, and two measure concepts, *length of stay* and *trajectory* are defined.

4.1 Database

The raw dataset collected from all sensors is huge and requires pre-processing for us to be able to extract meaningful information from it. These data basically consist of a unique encrypted identifier for every mobile device and two timestamps, which correspond to the first and last times such a device has been detected by the sensor. Then a database and a query engine were built to reorganize these data for our analyses (see table 2 for an example). Let us indicate some of the tags used for organizing and analysing the raw data. *Rffr* is the unique encrypted identifier. *Date* is the year, month and day when the data were collected. *Path* indicates the nodes that a mobile device has visited and it is represented by a sequence of node numbers, from 0 to 9, separated by a colon (":"). *Nodes* represents the total number of nodes that a device has visited during its whole trajectory, while *distinct nodes* indicates the number of different nodes which a mobile device has visited. *Checkin* is the moment when the signal of a mobile device is first detected in the museum (i.e. at the first node of the trajectory) and *checkout* is the moment when it disappears from the last node (i.e., when the device has left the museum). *Staylength* is the time difference between *Checkout* and *Checkin* and it represents the total duration of stay of a mobile device inside the museum.

Table 2. Example of the data set.

Rffr	Date	Path	Distinct nodes	Nodes	Checkin	Checkout	Staylength
Unique ID	2010-04-30	0:3:8:7:0	4	5	09:04:35	11:07:52	02:03:17

4.2 Collected sample

A high frequentation 10-day period in May 2010 was selected to perform a first analysis of visitor's behaviour. During this audit period, our installation recorded the presence of 12,944 unique devices. Through the data cleaning process we removed the logs from security and museum staff by looking at their recurrence, and the time of their presence (e.g. outside visiting times). Also, it was found out that the logs from two sensors had erroneous time synchronization and had to be discarded. Indeed, synchronization is a key element of our approach, for instance to infer the sequence of visit.

4.3 Measures definition

A sensor log reveals the visitor's presence at a node: once a Bluetooth-enabled mobile device enters the detectable area, the sensor continues to receive the signal emitted from the device until it disappears from its range. Each sensor records the first time the device appears as a check-in time at then records the time when the signal of the device disappears, as the checkout time. The difference between both time stamps is the length of the stay at the node. If the nodes visited are ordered by time, and then the checkin time at the first node in the trajectory and the checkout time at the last are selected, the values of the total duration of the visit to the museum will be obtained. As it can be seen, synchronization of sensors plays a key role for the collected data to be meaningful.

On the other hand when a unique Bluetooth identifier is logged out with time stamp (t1) by sensor A and some time later is logged in with time stamp (t2) by sensor B, the difference between t2 and t1 measures the travel time. The sequential movement of a mobile device detected by a pair of sensors (e.g. A-B) is defined as a trajectory with the travel time of t2-t1 minutes. The concept of trajectory in our research is different from that obtained with GPS systems, which indicate precise locations. The trajectories are obtained from Bluetooth detection through the time stamped sequential transition of a mobile device detected through different nodes (e.g. sequence of A-B-D), while GPS can describe the precise movement of the device. Our measures are, in this way, indirect ones.

5 Results

Using the Bluetooth data and concepts described previously, a novel approach is developed to analyse the spatial use in the Louvre museum. The following subsections present the on-going analysis efforts built around these concepts and the initial findings in order to obtain indicators for crowd management and to extract the frequent patterns in visitors' behaviour. In 5-1 the representativeness of the sample captured by our sensors is discussed. In 5-2 an analysis of the use of the Pyramid space related to the visitors' trajectories, which may reveal the distribution of visitors' presence and its basic flow in the museum, as all the visitors use the Pyramid as entrance and exit is presented. In 5-3 an analysis of visitors' trajectories and the time spent in each route is presented, revealing the existence of frequent patterns according to these two parameters.

5.1 Representativeness of the collected sample

Only a part of the visitors have got devices with Bluetooth, and only a part of them are enabled so that they are detectable. The number of devices detected at the entrance is compared with the official museum head counts and ticket sales to understand the representativeness of our Bluetooth data. The sample represented between 5.9% and 8.7% of the visitors with a strong, positive correlation of +80%, providing support for its representativeness. Figure 3 shows the linear regression fit of the numbers of detected devices and official counts of each day, with data of 101 days.

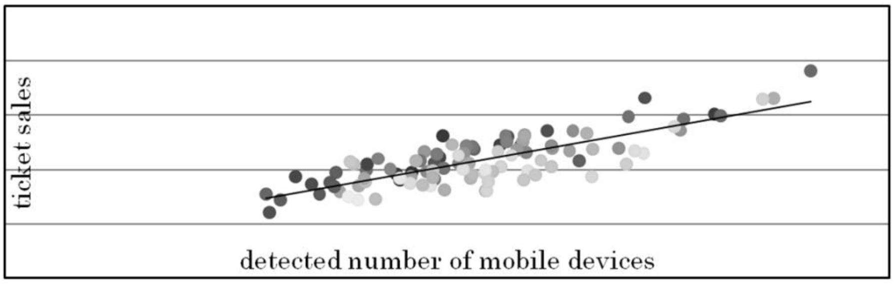


Fig. 3. Correlation between detected devices and visitors estimations per day

5.2 Use of the Pyramid space

The Pyramid space serves for distributing the visitors through three museum accesses, named *Denon*, *Sully* and *Richelieu*. As it is the starting point for almost all the museum visitors, it is important to identify and analyse their spatial use in order to devise more efficient and flexible policies for the museum. Since all the sensors are installed on the trails that lead to the Venus de Milo, along the Denon area, mobile devices detected by sensors in nodes 0 to 8 represent people who visited such area. These are the data we deal with in this section.

Sensor 0 is the Hall, and due to the museum's spatial layout, routes 0-3, 0-4, 0-7 and 0-8 mean that the *Denon access* has been used; routes 0-5, 0-6 and 0-9 mean use of either *Sully* or *Richelieu access*; route 0-0 means that only the Sully or Richelieu areas were visited (see Fig.4). Our data indicate that 76% of visitors used the *Denon access* while only 23% used either of the other two.

However, if one focuses on visitor's exit behaviour, i.e. moving towards node 0 or the Hall, the spatial use tendency changes. The most used route leading to node 0 was the 3-0 (25%), followed by the 7-0 (around 17%) and the 5-0 (around 16%); around 40% of the visitors left the museum through the *Sully* or *Richelieu* access, while 60% used *Denon* as their exit route, which means a decrease of an absolute 16% of the latter.

5.3 Visitor's trajectories

In this subsection, visitors' trajectories, their average length of stay and their relationships are analysed to discover frequent patterns or trends in visitors' behaviours. Sequential pattern mining (Agrawal & Srikant, 1995) has received much attention in the recent decade to find frequent sequences of events in data with a temporal component, with transition probabilities between events. However, extracting meaningful patterns requires appropriate algorithms and parameters. In the following, the grounding work and initial findings of our analysis are presented.

Most used trajectory and visitor's transition rates between nodes. Clarifying visitors' most used trajectory and their nodes transition rates helps to uncover hidden rules behind the seemingly disordered dataset. The data correspond to the route starting by 0-8, which is used by 60.6% of the visitors (7721 devices). Within the people who took this route, 71.1% moved to node 3 (0-8-3), while 13.5% went back to node 0 (0-8-0); 7.2% do not have any further records after node 8 (0-8), and it is assumed that they should have finished their itineraries without being registered by node 0 again. This is because if they would have continued visiting the museum, some of the other sensors should have detected them. Thus, 20.8% of these visitors (12.6% of the total) came to the museum just to take the 0-8-0 route. The popularity of this route might be due to the presence of two major works, Mona Lisa, located between nodes 7 and 8, and Winged Victory of Samothrace (in node 8). Moreover, the spatial structure of the museum strengthens the link between those two works, and thus an important sequential pattern including a visit to the Mona Lisa, followed by the Winged Victory of Samothrace, and then the Italian Gallery might exist. Let us perform a more detailed analysis considering the objects in the spatial structure, to reveal patterns.

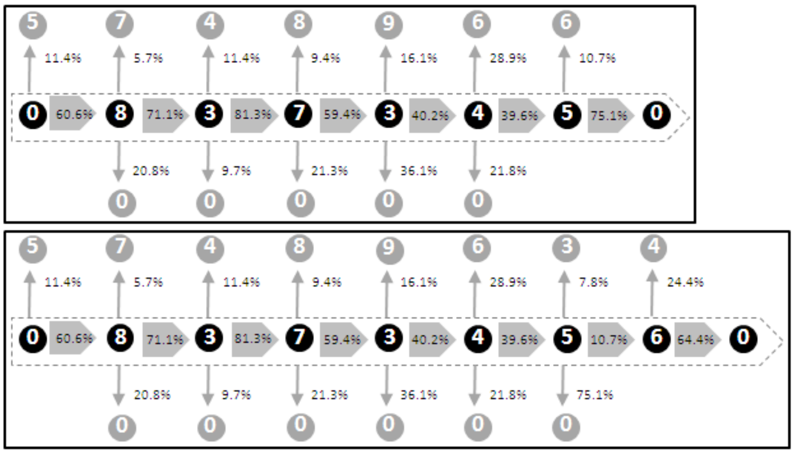


Fig. 5. Diagram of nodes, and percentages of visitors moving between them. Above, the most used trajectory is shown; below, the second most used one.

The visitors' distribution rate from each node to every subsequent node is iteratively computed until the route finished at 0. The percentages that appear with each arrow (Fig.5) are these transition rates from each node to the next ones. This makes it easier to understand quantitatively the visitors' flows. For instance, the upper part of fig.5 shows the most used trajectory (0-8-3-7-3-4-5-0) while the lower one shows the second one (0-8-3-7-3-4-5-6-0) both with thick arrows. Thin arrows express the second or third higher transition rates from a node to the following ones.

Various findings from the diagram can be extracted, but the most visible outcome is the strong connection between nodes 8 and 3, and between nodes 3 and 7. While 71.1% of visitors moved to node 3 after visiting node 8, 81.3% of them went to node 7 after visiting node 3. Concerning node 3, the analysis shows that people tend to use it to make a change of direction as it is in the same way from the Pyramid space. In a similar way as the Pyramid space, which distributes people for three accesses, node 3 also serves for distributing visitors to other places.

All of these findings demonstrate that our methodologies can reveal unknown aspects of visitors' spatial use, which observation and traditional interview-based approaches could not clarify at small scale in spatiotemporal terms.

Relationship between length of stay and number of visited nodes. Next, the relationship between the average stay length and the number of visited nodes of each trajectory is analysed, as it can provide another pattern of visitors' behaviour and spatial use in the museum.

Table 3. Average length of stay corresponding to each trajectory

Trajectory	**Average length of stay**

0-8-0	2:46:27
0-8-3-0	2:49:11
0-8-3-7-0	2:45:21
0-8-3-7-3-0	2:35:17
0-8-3-7-3-4-0	2:19:58
0-8-3-7-3-4-5-0 (most used trajectory)	2:19:14

The regression line of the number of nodes in a trajectory versus the average staying time has a negative slope, meaning that the larger the number of nodes in the trajectory, the shorter the visit to the museum lasts (with an r2 value of 0.85) – which opposes to the most obvious assumption.

6 Discussion, conclusions and future work

The paper shows, through different analyses, that Bluetooth data can throw new light on spatial use and visitors' behaviours at the building scale. Namely, evidence on the use of the different accesses as the entrance route by visitors and an exit pattern, which is different from the entrance pattern, has been given. This evidence would be difficult and costly to obtain from observations and surveys. Again, these traditional methods would have had difficulties to offer estimates of percentages of visitors that have followed different trajectories, to detect the importance of the 0-8-0 route, or to easily detect, from raw data, the role of node 3 as a crossroads. Another example of the power of the simple analysis on the data is the inverse relation found between length of stay and number of visited nodes.

These initial findings suggest that the methodology proposed has a great potential to clarify the features of the space and its use by visitors in small spatiotemporal scales with unprecedented accuracy. For example, comparing the results of audits of different periods would offer the possibility of obtaining results of seasonal nature; analysing audits of a large amount of data collected, finer detail of patterns and relationships could be obtained – beyond the crude relationship of average visit length and number of nodes. Collecting audits during longer periods requires only very small extra empirical effort besides the one already described; and the analysis would only mean extending and refining the analytical tools.

Before discussing more details of our current work, let discuss some aspects of the data obtained. As seen, the data correspond to a small sample of the visitors – although very large compared to the typical sample used in surveys, and without subjective bias -, but seems to be reasonably valid in terms of the correlation shown. However, more work should be done to clarify the extent to which the sample is representative, as carrying a Bluetooth device set as enabled might be a significant bias. With respect to other data collection strategies involving users carrying specific devices and consequently being aware of them, data appear to be free from potential bias, and the dataset obtained is larger.

Secondly, the data obtained are usually noisy; however, examples of strategies for checking data consistency from the data themselves, and for cleaning it have been given. Larger audits, which would offer larger datasets, can help to strengthen this aspect.

Let turn now to current and future work. Exploring some of the aspects mentioned before based on larger audits of the Louvre has started. And based on the current results, indicators for crowd management and an algorithm for sequential mining are being developed to discover frequent patterns and the underlying association rules. The dynamic estimation of the density and flow of visitors in and between nodes could be associated with the indicator of the relation between pedestrian flow and its density (Seyfried, Steffen, Klingsch, & Boltes, 2005) for more dynamic crowd management. While several attempts have been made to extract meaningful frequent trajectory patterns and predict further movements of objects at a variety of scales from region and city (Giannotti, Nanni, Pedreschi, & Pinelli, 2007) to retail shop (Larson, Bradlow, & Fader, 2005), improved mining techniques and parameter settings depending on the nature of the data would be needed in order to achieve these goals. Explore similar pedestrian data collected in unconstrained environments has started, and it should help to substantiate the previous statement.

As a final point, one should remark that the understanding of the patterns in visitors' behaviour and its prediction will enable to optimize the spatial layout of objects, human resources and facilities, including advertising and visitor information points, to respond to the increasing tourism demand. It could become a strong management tool not only for museums but also for urban environments in the tourism flourishing age.

References

Agrawal, R. & Srikant, R. (1995). Mining sequential patterns. In S. Yu. Philip, Arbee L.P. Chen (Eds.), *ICDE '95: Proceedings of the Eleventh International Conference on Data Engineering*, Washington DC. IEEE Computer Society.

Ahas, R., Aasa, A., Roose, A, Mark, U. & Silm, S. (2008) Evaluating passive mobile positioning data for tourism surveys: An Estonian case study. *Tourism Management* 29(3): 469-486.

Antonini, G., Bierlaire, M. & Weber, M. (2006). Discrete choice models of pedestrian walking behaviour. *Transportation Research Part B Methodological* 40(8): 667-687.

Asakura, Y. & Iryob, T. (2007). Analysis of tourist behaviour based on the tracking data collected using a mobile communication instrument. *Transportation Research Part A: Policy and Practice* 41(7): 684-690.

Barceló, J., Montero, L., Marqués, L. & Carmona, C. (2010). Travel Time Forecasting and Dynamic Origin-Destination Estimation for Freeways Based on Bluetooth Traffic Monitoring. Transportation Research Record: Journal of the Transportation Research Board 2175: 19-27.

Eagle, N. & Pentland, A. (2005) Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing* 10(4):255-268.

Giannotti, F., Nanni, M., Pedreschi, D. & Pinelli, F. (2007). Trajectory Pattern Mining. *Proceedings of the 13th ACM SIGKDD international conference on knowledge discovery and data mining KDD 07*, San Jose: 330-339.

Girardin, F., Calabrese, F., Dal Fiore, F., Ratti, C. & Blat, J. (2008). Digital footprinting: Uncovering tourists with user-generated content. *Pervasive Computing, IEEE* 7(4): 36-43.

Girardin, F., Dal Fiore, F., Ratti, C., & Blat, J. (2008). Leveraging explicitly disclosed location information to understand tourist dynamics: a case study. *Journal of Location Based Services* 2(1): 41-56.

Girardin, F., Dillenbourg, P. & Nova, N. (2009). Detecting air travel to survey passengers on a worldwide scale. *Journal of Location Based Services* 3(3): 210-226.

Hillier, B. & Tzortzi, K. (2006). Space Syntax: The Language of Museum Space. In S. MacDonald, (Ed.), A Companion to Museum Studies, London, Blackwell Publishing.

Hooper-Greenhill, E. (2006). Studying visitors. In S. MacDonald (Ed.), *A Companion to Museum Studies*, London, Blackwell Publishing.

Kanda, T., Shiomi, M., Perrin, L., Nomura,T., Ishiguro, H. & Hagita, N. (2007). Analysis of people trajectories with ubiquitous sensors in a science museum. *Proceedings 2007 IEEE International Conference on Robotics and Automation (ICRA'07)*: 4846-4853.

Kostakos, V., O'Neill, E., Penn, A., Roussos, G. & Papadongonas, D. (2010). Brief encounters: sensing, modelling and visualizing urban mobility and copresence networks. *ACM Transactions on Computer Human Interaction* 17(1): 1-38.

Larson, J.S., Bradlow, E.T. & Fader, P.S. (2005). An exploratory look at supermarket shopping paths. *International Journal of Research in Marketing*, 22: 395-414.

Nicolai, T., Yoneki, E., Behrens, N. & Kenn, H. (2006). Exploring social context with the Wireless Rope. In R. Meersman, Z. Tari, & P. Herrero, (Eds.), *On the move to meaningful internet systems 2006: LNCS, vol 4277,* Heidelberg: Springer.

O'Neill, E., Kostakos, V., Kindberg, T., Fatah gen. Schieck, A., Penn, A., Stanton Fraser,D. & Jones,T. (2006). Instrumenting the city: Developing methods for observing and understanding the digital cityscape. In P. Dourish and A. Friday, (Eds.), Ubicomp 2006, LNCS 4206: 315-332.

Paulos, E. & Goodman, E. (2004). The familiar stranger: anxiety, comfort, and play in public places. *Proceedings of the SIGCHI conference on Human factors in computing systems* 6(1): 223-230.

Pereira, F. C., Vaccari, A., Giardin, F., Chiu, C., & Ratti, C. (2012) Crowdsensing in the web: analysing the citizen experience in the urban space. In M. Foth, L. Forlano, C. Satchell, and M. Gibbs (Eds), *From Social Butterfly to Engaged Citizen: Urban Informatics, Social Media, Ubiquitous Computing, and Mobile Technology to Support Citizen Engagement*, Cambridge (MA): MIT Press.

Ratti, C., Pulselli, R., Williams, S. & Frenchman, D. (2006). Mobile Landscapes: using location data from cell phones for urban analysis. *Environment and Planning B: Planning and Design* 33(5): 727-748.

Sanfeliu, A., Llácer, M.R., Gramunt, M.D., Punsola, A. & Yoshimura, Y. (2010). Influence of the privacy issue in the Deployment and Design of Networking Robots in European Urban Areas. *Advanced Robotics* 24(13): 1873-1899.

Seyfried, A., Steffen, B., Klingsch, W. & Boltes, M. (2005). The fundamental diagram of pedestrian movement revisited. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(10): 10002-10002.

Yalowitz, S. S. & Bronnenkant, Kerry. (2009) Timing and Tracking: Unlocking Visitor Behavior. *Visitor Studies* 12(1): 47-64.

Acknowledgements

We would like to thank the Studies and Research Department of the Louvre Museum as well as BitCarrier for their support. We are indebted to many people at Universitat Pompeu Fabra and Lift Lab for the stimulating research environments and their generous feedback. Of course, any shortcomings are our sole responsibility.