Incorporating Mobility Patterns in Pedestrian Quantity Estimation and Sensor Placement

Thomas Liebig, Zhao Xu, and Michael May

Fraunhofer IAIS
Schloss Birlinghoven, 53754 Sankt Augustin, Germany
{Thomas.Liebig,Zhao.Xu,Michael.May}@iais.fraunhofer.de

Abstract. Pedestrian quantity estimation receives increasing attention and has important applications, e.g. in location evaluation and risk analysis. In this work, we focus on pedestrian quantity estimation for event monitoring. We address the problem (1) how to estimate quantities for unmeasured locations, and (2) where to place a bounded number of sensors during different phases of a soccer match. Pedestrian movement is no random walk and therefore characteristic traffic patterns occur in the data. This work utilizes traffic pattern information and incorporates it in a Gaussian process regression based approach. The empirical analysis on real world data collected with Bluetooth tracking technology during a soccer event at Stade des Costières in Nîmes (France) demonstrates the benefits of our approach.

Keywords: Pedestrian Quantity Estimation, Trajectory, Gaussian Process Regression, Graph Kernels, Sensor Placement

1 Introduction

Major public events such as soccer matches, concerts and festivals attract thousands or even millions of visitors. On the one hand this offers interesting business opportunities for event organizers, advertisement companies and street marketers. On the other hand it also creates a growing financial risk for the organizers due to huge expenses, and safety risks for the guests themselves. Understanding movement behaviour and identification of attractors and distractors gives insights on visitor preferences and motivations during a particular event. This can help in avoiding risks by better management of visitor flows. Various locations and attractions can be ranked by their popularity, safety or frequency, and measures against over-crowding can be taken immediately or for future events.

Sensor technologies that are currently in use to measure people quantities automatically are surveys, video surveillances, GPS, and Bluetooth scanners. Whereas the first solution (surveys) is expensive and hardly representative due to the non-random sampling among all visitors, the second one (video surveillance) depends on weather, brightness and density of the people and does sometimes require special scaffoldings to carry the cameras. GPS finally is not available everywhere, e.g. indoors and in urban canyons. In this paper we perform our tests

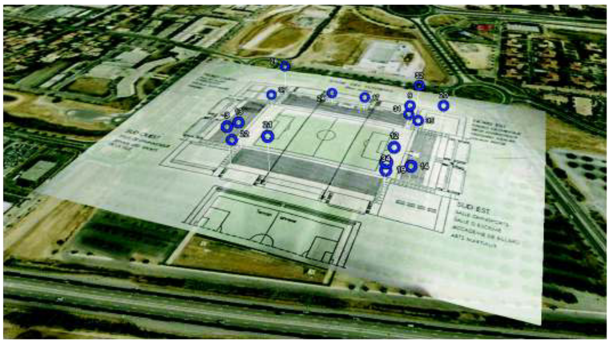


Fig. 1. 3D Sensor Placement at Stade des Costières, Nîmes (France) 05/08/2011.

on a Bluetooth tracking dataset collected during a soccer match at the Stade des Costières, Nîmes (France) [1]. The data was collected using 17 Bluetooth beacons [2] at various locations in the stadium (Figure 1).

This work addresses the question where a fixed number of automatic pedestrian quantity sensors (i.e. Bluetooth beacons) is to be located during a mass event in order to get an adequate estimate on the movement of the visitors within the site. Our approach addresses the following questions:

– How can pedestrian quantities be estimated from a relatively small number of empirical measurements?
– At which places should a constrained number of pedestrian quantity sensors be located?

Often, available data for investigating these questions is limited to a small number of measurements and some prior knowledge, e.g., floor plan sketches or knowledge on preferred routes by local domain experts. Incorporating prior knowledge is thus essential to address the above challenges. However, so far there are few approaches that explicitly take into account the movement patterns, although pedestrians generally show some move preferences [3–6], especially in closed environments, e.g., sport stadiums.

In this paper we address both, pedestrian quantity estimation and sensor placement in the case where movement patterns are provided as background knowledge (Section 3) and the acquisition of movement patterns from Bluetooth observation data (Section 4). The paper is structured as follows. Section 2 discusses related work and gives an introduction to episodic movement data and its analysis. In Section 3 we introduce our Bayesian method (Gaussian processes). Section 4 highlights the application to the real world dataset. We conclude in Section 5.

2 Related Work

Bluetooth monitoring has found a number of interesting applications in recent years. Besides event monitoring, also other successful applications of Bluetooth tracking technology are described in the literature. In [7] various scanners were placed at Dutch train stations to record transit travellers. Accurate location and tracking of objects within complex facilities is another important research topic [8]. Bluetooth tracking is also used to monitor a sample of visitors [1, 8, 9] and extract their route choices [1, 10]. The work presented in [11] uses Bluetooth tracking to record people in a public transportation network, whereas [12] gives a general overview on possibilities using Bluetooth tracking technology. In a few works time-geography and movement patterns are addressed as well [9, 13].

Bluetooth tracking is based on collecting episodic movement data (EMD) [9]. In GPS-less environments episodic movement data is the major representation of pedestrian mobility. Differently from outdoor pedestrian quantity estimation, continuous tracking technologies such as GPS cannot be used in many closed environments due to the lack of a GPS signal in buildings and/or expensive deployment of the hardware. Instead, recently developed alternative technologies such as light beams, video surveillance, and Bluetooth meshes record episodic movement data or its location-based-aggregate, presence counts, at low expenses. Episodic movement data is represented by tuples $< o, p, t >$ of moving object identifier o, discrete location identifier p and a time stamp t. The location-based-aggregate, presence counts, for time interval Δt, is also known as *number of visits*, *quantity* or *traffic frequency*. It is defined as $NV(p, \Delta t) = | < o, p, t >, t \in \Delta t|$. The *number of moves* among two locations p_i and p_j is similarly defined as $NM(p_i, p_j, \Delta t) = | < o, p_i, p_j, t >, t \in \Delta t|$. Other prominent examples of episodic movement data are spatio-temporal activity logs, geo-tagged photos, cell based tracking data and billing records.

Episodic movement data poses great challenges for existing data mining algorithms based on (linear) interpolation between data points. For example, speed and movement direction cannot be directly derived from episodic data; trajectories may not be depicted as a continuous line; and densities cannot directly be computed. The reason is that there are normally unmeasured locations between two measurements that cannot be reliably inferred by linear or other parametric interpolation.

Though this data is thus difficult to use for individual movement or path analysis, it still contains rich information on group movement on a coarser level. Our approach is to aggregate movement in order to overcome some of the uncertainties present at the individual level. Deriving the number of objects for spatio-temporal areas and transitions among them gives interesting insights on spatio-temporal behavior of moving objects. As a next step to support analysts, [9] proposes clustering of the spatio-temporal presence and flow situations (see Figure 3). In this figure the colour shading, which supports a visual understanding and analysis of the flows, results from Sammon‘s mapping [14]. To be more precise, the two-dimensional clustering of the flow situations (vector among all sensors) is mapped on a colour plane. As a result, similar flows get similar colours, and difference between flows corresponds to difference between colours. The different stages of the match are visible, and are subject for data partition in Section 4.

3 Pedestrian Quantity Estimation with Movement Patterns

Although pedestrians show systematic behavior and move preferences, especially in closed environments, e.g., stadiums, concert halls or trade fairs, few approaches systematically take into account the trajectory patterns for analysis. However, incorporating prior knowledge on pedestrian movement is essential to address the two questions posed above (see section 1). Existing traffic volume estimation methods, e.g., k-nearest neighbour [15, 16] and standard Gaussian process regression [17], do not take into account this form of expert knowledge and thus may not effectively provide accurate estimations, e.g. in case of side corridors.

To estimate the traffic volume at unmeasured locations, we propose in [18] a nonparametric Bayesian method, Gaussian Processes (GP) with a random-walk based trajectory kernel. The method explores not only the commonly used information known from the literature, e.g. traffic network structures and recorded presence counts NV at some measurement locations, but also the move preferences of pedestrians (trajectory patterns) collected from the sensors. As firstly introduced in [18], we provide here a brief discussion on the GP approach for quantity estimation and sensor placement. Consider a traffic network $\mathcal{G}(\mathbf{V}, \mathbf{E})$ with N vertices and M edges. For some of the edges, we observe the pedestrian quantities, denoted as $\mathbf{y} = \{y_s := NV(\tilde{v}_s, \Delta t) : s = 1, \ldots, S\}$. Additionally, we have information about the major pedestrian movement patterns $\mathcal{T} = \{T_1, T_2, \ldots\}$ over the traffic network, collected from the local experts or the tracking technology (e.g. Bluetooth). The pedestrian quantity estimation over traffic networks can be viewed as a link prediction problem, where the predicted quantities associated with links (vertices) are continuous variables.

In the literature on statistical relational learning [19, 20], a commonly used GP relational method is to introduce a latent variable to each vertex, and to model the values of edges as a function of latent variables of the involved vertices, e.g. [21, 22]. Although these methods have the advantage that the problem size remains linear in the size of the vertices, it is difficult to find appropriate functions to encode the relationship between the variables of vertices and edges for different applications.

The observed pedestrian quantities (within a time interval Δt) are conditioned on the latent function values with Gaussian noise ϵ_i: $y_i = f_i + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2)$. As mathematical form and parameters of the function are random and unknown, f_i is also unknown and random. For an infinite number of vertices, the function values $\{f_1, f_2, \ldots\}$ can be represented as an infinite dimensional vector. Within a nonparametric Bayesian framework, we assume that the infinite dimensional random vector follows a Gaussian process (GP) prior with mean function

$m(x_i)$ and covariance function $k(x_i, x_j)$ [23]. In turn, any finite set of function values $\mathbf{f} = \{f_i : i = 1, ..., N\}$ has a multivariate Gaussian distribution with mean and covariances computed with the mean and covariance functions of the GP [23]. Without loss of generality, we assume zero mean so that the GP is completely specified by the covariance function. Formally, the multivariate Gaussian prior distribution of the function values $\mathbf{f}$ is written as $P(\mathbf{f}|\mathbf{X}) = \mathcal{N}(0, K)$, where K denotes the $N \times N$ covariance matrix, whose ij-th entry is computed in terms of the covariance function. If there are vertex features $\mathbf{x} = \{x_1, ..., x_N\}$ available, e.g., the spatial representation of traffic edges, a typical choice for the covariance function is the squared exponential kernel with isotropic distance measure.

Since the latent variables $\mathbf{f}$ are linked together into an edge graph $\mathcal{G}$, it is obvious that the covariances are closely related to the network structure: the variables are highly correlated if they are adjacent in $\mathcal{G}$, and vice versa. Therefore we can also employ graph kernels, e.g. the regularized Laplacian kernel, as the covariance functions:

$$K = \left[\beta(L + I/\alpha^2)\right]^{-1}, \tag{1}$$

where α and β are hyperparameters. L denotes the combinatorial Laplacian, which is computed as $L = D - A$, where A denotes the adjacency matrix of the graph $\mathcal{G}$. D is a diagonal matrix with entries $d_{i,i} = \sum_j A_{i,j}$.

Although graph kernels have some successful applications to public transportation networks [17], there are probably limitations when applying the network-based kernels to the scenario of closed environments: the pedestrians in a train station or a shopping mall have favorite or commonly used routes, they are not randomly distributed on the networks. In a train station, the pedestrian flow on the main corridor is most likely unrelated to that on the corridors leading to the offices, even if the corridors are adjacent. To incorporate the information of the move preferences (trajectory patterns, collected from the local experts or tracking technology) into the model, we explore a graph kernel inspired with the diffusion process [24]. Assume that a pedestrian randomly moves on the edge graph $\mathcal{G}$. From a vertex i he jumps to a vertex j with $n_{i,j}^k$ possible random walks of length k, where $n_{i,j}^k$ is equal to $[A^k]_{i,j}$. Intuitively, the similarity of two vertices is related to the number and the length of the random walks between them. Based on diffusion process, the similarity between vertices v_i and v_j is defined as

$$s(v_i, v_j) = \left[\sum_{k=1}^{\infty} \frac{\lambda^k}{k!} A^k\right]_{ij}, \tag{2}$$

where $0 \leq \lambda \leq 1$ is a hyperparameter. All possible random walks between v_i and v_j are taken into account in similarity computation, however the contributions of longer walks are discounted with a coefficient $\lambda^k/k!$. The similarity matrix is not always positive semi-definite. To get a valid kernel, the combinatorial Laplacian is used and the covariance matrix is defined as [24]:

$$K = \left[\sum_{k=1}^{\infty} \frac{\lambda^k}{k!} L^k\right] = \exp(\lambda L) . \tag{3}$$

On a traffic network within closed environment, the pedestrian will move not randomly, but with respect to a set of trajectory patterns and subpatterns denoted as sequences of vertices, e.g.,

$$\left\{ \begin{array}{l} T_1 = v_1 \to v_3 \to v_5 \to v_6, \\ T_2 = v_2 \to v_3 \to v_4, \\ \quad \dots \end{array} \right\}. \tag{4}$$

Each trajectory pattern T_ℓ can also be represented as an adjacency matrix in which $\hat{A}_{i,j} = 1$ iff $v_i \to v_j \in T_\ell$ or $v_i \leftarrow v_j \in T_\ell$. The subpatterns are subsequences of the trajectories. For example, the subpatterns of T_1 are $\{v_1 \to v_3, v_3 \to v_5, v_5 \to v_6, v_1 \to v_3 \to v_5, v_3 \to v_5 \to v_6\}$. Given a set of trajectory patterns $\mathcal{T} = \{T_1, T_2, \dots\}$, a random walk is valid and can be counted in similarity computation, if and only if all steps in the walk belong to $\mathcal{T}$ and subpatterns of $\mathcal{T}$. Thus we have

$$\begin{aligned} \hat{s}(v_i, v_j) &= \left[\sum_{k=1}^{\infty} \frac{\lambda^k}{k!} \hat{A}^k \right]_{ij}, & \hat{K} &= \left[\sum_{k=1}^{\infty} \frac{\lambda^k}{k!} \hat{L}^k \right] = \exp(\lambda \hat{L}) \\ \hat{A} &= \sum_\ell \hat{A}_\ell, & \hat{L} &= \hat{D} - \hat{A}, \end{aligned} \tag{5}$$

where $\hat{D}$ is a diagonal matrix with entries $\hat{d}_{i,i} = \sum_j \hat{A}_{i,j}$.

For pedestrian quantities $\mathbf{f}_u$ at unmeasured locations u, the predictive distribution can be computed as follows. Based on the property of GP, the observed and unobserved quantities $(\mathbf{y}, \mathbf{f}_u)^T$ follows a Gaussian distribution

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_u \end{bmatrix} \sim \mathcal{N} \left(0, \begin{bmatrix} \hat{K}_{\overline{u},\overline{u}} + \sigma^2 I & \hat{K}_{\overline{u},u} \\ \hat{K}_{u,\overline{u}} & \hat{K}_{u,u} \end{bmatrix} \right), \tag{6}$$

where $\hat{K}_{u,\overline{u}}$ is the corresponding entries of $\hat{K}$ between the unmeasured vertices u and measured ones $\overline{u}$. $\hat{K}_{\overline{u},\overline{u}}$, $\hat{K}_{u,u}$, and $\hat{K}_{\overline{u},u}$ are defined equivalently. I is an identity matrix of size $|\overline{u}|$. Finally the conditional distribution of the unobserved pedestrian quantities is still Gaussian with the mean m and the covariance matrix Σ:

$$\begin{aligned} m &= \hat{K}_{u,\overline{u}} (\hat{K}_{\overline{u},\overline{u}} + \sigma^2 I)^{-1} \, \mathbf{y} \\ \Sigma &= \hat{K}_{u,u} - \hat{K}_{u,\overline{u}} (\hat{K}_{\overline{u},\overline{u}} + \sigma^2 I)^{-1} \, \hat{K}_{\overline{u},u} \ . \end{aligned}$$

Besides pedestrian quantity estimation, incorporating trajectory patterns also enables effectively finding sensor placements that are most informative for traffic estimation on the whole network. To identify the most informative locations $\mathcal{I}$, we employ the exploration strategy, maximizing mutual information [25]

$$\arg\max_{\mathcal{I} \subset \mathbf{V}} H(\mathbf{V} \backslash \mathcal{I}) - H(\mathbf{V} \backslash \mathcal{I} \mid \mathcal{I}) \ . \tag{7}$$

It is equal to finding a set of vertices $\mathcal{I}$ which maximally reduces the entropy of the traffic at the unmeasured locations $\mathbf{V}\backslash\mathcal{I}$. Since the entropy and the conditional entropy of Gaussian variables can be completely specified with covariances, the selection procedure is only based on covariances of vertices, and does not involve any pedestrian quantity observations. To solve the optimization problem, we employ a poly-time approximate method [25]. In particular, starting from an empty set $\mathcal{I} = \emptyset$, each vertex is selected with the criterion:

$$v_* \leftarrow \arg\max_{v \in \mathbf{V}\backslash\mathcal{I}} \; H_\epsilon(v \mid \mathcal{I}) - H_\epsilon(v \mid \overline{\mathcal{I}}) \; , \tag{8}$$

where $\overline{\mathcal{I}}$ denotes the vertex set $\mathbf{V}\backslash(\mathcal{I} \cup v)$. $H_\epsilon(x|Z) := H(x|Z')$ denotes an approximation of the entropy $H(x|Z)$, where any element z in $Z' \subset Z$ satisfies the constraint that the covariance between z and x is larger than a small value ϵ. Within the GP framework, the approximate entropy $H_\epsilon(x|Z)$ is computed as

$$\begin{aligned} H_\epsilon(x \mid Z) &= \frac{1}{2} \ln 2\pi e \sigma^2_{x|Z'} \\ \sigma^2_{x|Z'} &= \hat{K}_{x,x} - \hat{K}^T_{x,Z'} \hat{K}^{-1}_{Z',Z'} \hat{K}_{x,Z'} \; . \end{aligned} \tag{9}$$

The term $\hat{K}_{x,Z'}$ is the corresponding entries of $\hat{K}$ between the vertex x and a set of vertices Z'. $\hat{K}_{x,x}$ and $\hat{K}_{Z',Z'}$ are defined equivalently. Given the informative trajectory pattern kernel, the pedestrian quantity observations at the vertices selected with the criterion (8) can well estimate the situation of the whole network. Refer to [18] for more details.

4 Real World Application

In this section, we test our approach on a dataset collected through Bluetooth tracking technology [1]. The analysis is inspired by the workflow presented in [13]. Instead of applying two phases we conduct our experiment in three consecutive phases.

- The *field study phase* is performed during (1) survey design and (2) data collection.
- The second *visual analysis phase* is conducted within the (3) data preparation, aggregation and visual analysis.
- In the *knowledge discovery phase* we conduct the (4) data mining step.

Next, each of the steps is described and experiments to the previously described sensor placement strategy are performed.

4.1 Field Study Phase

For data collection a mesh of 17 Bluetooth sensors has been deployed within a soccer stadium (Stade des Costières, Nîmes at France) during a soccer match on

05.08.2011. The three-dimensional sensor placement is depicted in Figure 1. All Bluetooth enabled devices (e.g. smartphones or intercoms) that pass at one of the sensors (more precisely its footprint) trigger the creation of a datalog entry consisting of the timestamp, the sensor identifier (which denotes the position), the radio signal strength and a hashed identifier for this particular device [26].

The range of the sensors is approximately 15 meters, thus there remain unobserved regions in the stadium as well as overlapping areas. Whenever a Bluetooth enabled (i.e. visible) mobile device traverses multiple sensors, it becomes re-detected. In this way, transition times as well as movement patterns can be reconstructed. However, the recorded data is episodic (see Section 2 for specifics on *Episodic Movement Data*) as it provides uncertainties on continuity, accuracy as well as coverage [9].

We recorded 47,589 data points from 553 different devices at 17 distinct locations. The average number of distinct visited sensor locations is 4.37, the median number is 2. The recorded movements have an average duration of 3 hours and 25 minutes. In total, about 14 percent of the visitors, 553 of 3898 (this official visitor number does not contain the people which worked there), have been recorded during the period of the match; thus we expect the dataset to be large enough to allow inferences from the sample to the whole population even for less frequent flows.

4.2 Visual Analysis Phase

The recorded Bluetooth tracking dataset contains sequence movement patterns, which can for example be extracted using the Teiresias algorithm [27], which was firstly applied to episodic movement data in [28]. Application of the algorithm reveals that the most frequent pattern with more than one location starts at the main entrance and ends at a tribune (depicted in Figure 2A). The movement in the stadium thus is not a random walk but aims at a target. These individual movement preferences cause correlations among the sensor readings. Next, we visually explore the correlations contained in the soccer dataset [1]. The visual analysis of movement dependencies among discrete regions is subject of our previous work presented in [29, 29]. There, the contained dependencies are represented by a Spatial Bayesian Network which connects the different regions by directed edges and associated conditional probability tables. In result, queries for co-visits of spatial regions given arbitrary (positive or negative) evidences can be answered. Next, we apply this method [30] to the presented dataset and study the contained movement preferences in detail (Figure 2).

For visualization of the three-dimensional dependencies, we created a Voronoi Dirichlet tessellation of a three-dimensional stadium model. Materials to the resulting geometries (colour and opacity) are assigned according to the probability distribution computed by the Spatial Bayesian Network. Figure 2 depicts the results of the Spatial Bayesian Network for four different queries. Red colours indicate a high visit probability; blue colours indicate a low probability. The yellow arrows in the picture mark the points of positive evidence. The picture A (in the upper-left corner) depicts the probability distribution given the evidence that the sensor at the ground floor (sensor 34 for comparison with Figure 1) has been visited. It is remarkable that the probability on this side of the stadium is high and low in most of the other parts. The places in the other tribunes (at the bottom of the pictures) that possess a relative high probability as well as the VIP rooms and thus visited by the catering staff and prominent visitors from all tribunes after the match ended.

In the next step we examine the impact of the staff and prominent guests by change of evidence to a restricted entry within the Spatial Bayesian Network. Results are depicted in picture B. All paths that have been used by the catering crew and safety deputies are inked in red which denotes a high probability of movement. The shops possess a relatively high probability. They were located in the uppermost floor of the two towers in the left side of the picture and also in the VIP lounges. Safety deputies helped us during data collection, thus it can be seen to the right that they visited sensor location three (top of the upper left tower, compare Figure 2) in order to check its presence. In the bottom of Figure 2 we combine multiple points of evidence within the query. To the left (picture C) is a visualization of the combined probability of the visitors at the entry to the major tribune and to the VIP entry. The visitors selected by this query distribute among the major tribune and within the VIP rooms. By further addition of evidence at sensor location three, the places considered so far reach their highest conditional probability. Most likely this untypical movement pattern depicted in picture D was our movement for maintenance of the sensors. The tribune to the left shows a very low probability as it could not be traversed. The tribune on the right was open for traversing before the match began. Thus, our visual analysis reflects these circumstances and helps to understand movement behavior contained in the dataset.

After visual analysis of the recorded spatial movement correlations our further visual analysis focuses on the temporal analysis and the preparation of the dataset for the data mining (i.e. sensor placement step). Since episodic movement data contains uncertainties on individual movement, the proposed approach in [9]

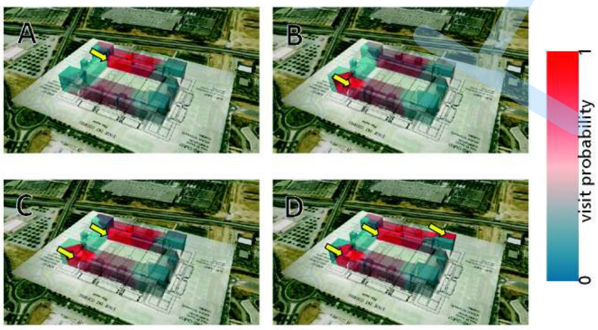
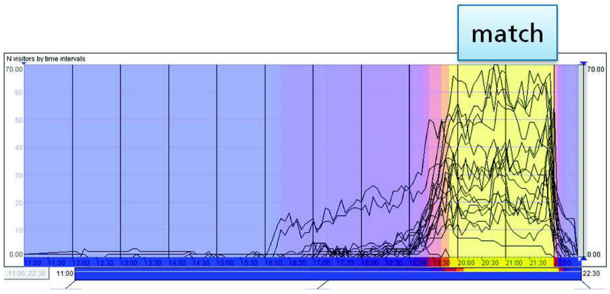
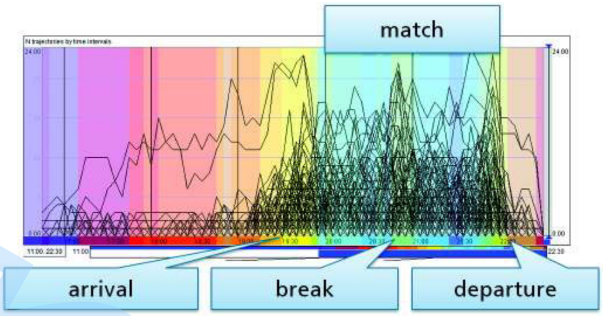


Fig. 2. Visual representation of the spatial correlations in the soccer dataset, yellow arrow denotes the evidence of the query



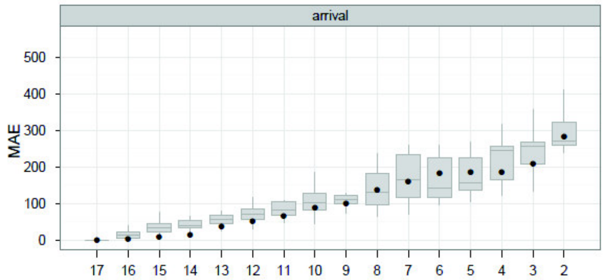
(a) Clustering of presence situations over time



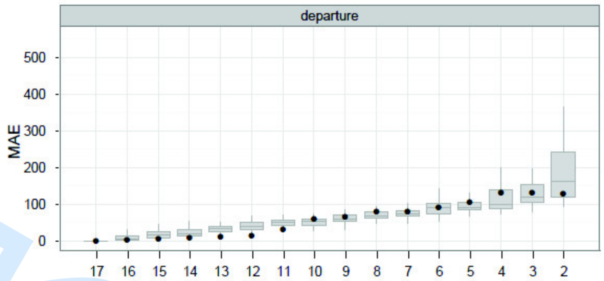
(b) Clustering of flow situations over time

Fig. 3. Temporal analysis of presence and flow situations.

is the spatio-temporal aggregation of *presence* and *moves*. This results in presence and flow situations which denote for a time interval Δt the total number of *visits* for each discrete location as well as the total number of *moves* among pairs of locations. Thus, in contrast to the existing workflow for Bluetooth tracking data analysis, the soccer dataset [1] is divided into three consecutive time intervals (arrival, match, departure) derived from the clustering of presence and flows (Figure 3). In this picture the lines represent the number of persons per scanner (Figure 3a) or the numbers of persons per link among two locations (Figure 3b). The background colouring of the Figure utilizes Sammons mapping [14] and was discussed in Section 2. Based on the achieved visual analysis of the flow data (depicted in Figure 3) the time-stamps for splitting are (14:00, 20:00, 21:45, 22:00). These time intervals correspond to the three different consecutive phases of the match: arrival of the visitors, match and the departure after the match. Note that in Figure 3b (which analyses the moves of the visitors) even the break of the match is visible. Movement of the stadium visitors differs in each of these time spans from its successive time interval (indicated by different colours in Figure 3b). During the match there is very low movement of the visitors. Thus, we perform our sensor placement experiments for the safety critical phases of *arrival* and *departure*.



(a) arrival dataset



(b) departure dataset

Fig. 4. MAE for random (grey boxplots) and trajectory pattern kernel based sensor placement (black dots) for different number of sensors

4.3 Knowledge Discovery Phase

The Gaussian process based sensor placement algorithm (Section 3) is applied to the two previously separated datasets (arrival and departure of the visitors). Thus, the recorded movement sequences of the visitors (studied in Figure 2) are considered as movement patterns. All of the recorded patterns are treated equally (we remove duplicates) without any weighting. The recorded *counts of visits* per sensor are subject for quantity estimation. This is also an important difference from our work presented in [18] where we model *counts of flows*.

The quantity estimation is performed with different numbers of sensors, starting from 17 up to 2. In each test we apply our sensor placement algorithm among the predefined locations chosen in the given dataset. The performance of the placement is then compared to random sensor placement (run 35 times each). The quantity estimation error is measured in mean absolute error MAE. Figure 4 depicts the performance for different numbers of sensors. The placement of 17 sensors (to the left) equals to the case were all 17 previously placed sensors (contained in the dataset) are used. In the next step, one of the sensors is omitted. The grey boxplot denotes performance for its random selection, the black dot the performance of our kernel based placement strategy.

The tests show that when omitting up to 6 of the applied sensors (35%) in the sensor mesh, our placement still outperforms random placement and has an acceptable absolute prediction error of 80 persons (2% of the total number of 3,898 visitors[1]).

5 Conclusion and Summary

The paper addressed the visitor quantity estimation in an event monitoring scenario under constraints (i.e., a bounded number of sensors). Thus we tackled the following two challenges (1) pedestrian quantity estimation from a relatively small number of empirical measurements, and (2) placement of the constrained number of quantity sensors. We proposed a novel method to determine where a fixed number of automatic pedestrian quantity sensors is to be located during a mass event in order to get an adequate estimate on the presence of the visitors within the site. Note that we considered here *counts of presence*, instead of *counts of moves*, which is subject to [18].

Our proposed method incorporates trajectory patterns for automatic sensor placement and quantity estimation. Real world experiments at a soccer stadium dataset show that our method holds potential for automatically determined sensor number reduction.

Future work may focus on reduction of communication costs among the sensor network, inclusion of mobile sensors (e.g. mobile Bluetooth sensors [31]) and creation of a dynamic pedestrian model.

Acknowledgments. The work was supported by the European Project Emergency Support System (ESS 217951) and the Fraunhofer ATTRACT Fellowship STREAM.

References

1. Liebig, T., Kemloh Wagoum, A.U.: Modelling microscopic pedestrian mobility using bluetooth. In: Proc. of the Fourth International Conference on Agents and Artificial Intelligience - ICAART'12, SciTePress (2012) 270–275
2. Bruno, R., Delmastro, F.: Design and Analysis of a Bluetooth-based Indoor Localization System. In: 8th International Conference on Personal Wireless Communications. (2003) 711–725
3. Liebig, T., Xu, Z.: Pedestrian monitoring system for indoor billboard evaluation. Journal of Applied Operational Research **4**(1) (2012) 28–36
4. Liebig, T.: A general pedestrian movement model for the evaluation of mixed indoor-outdoor poster campaigns. In: Proc. of the Third International Conference on Applied Operation Research - ICAOR'11, Tadbir Operational Research Group Ltd. (2011) 289–300

[1] info to the match at http://www.foot-national.com/match-foot-nimesvannes-32912.html, last accessed 08/05/2012

5. Liebig, T., Stange, H., Hecker, D., May, M., Körner, C., Hofmann, U.: A general pedestrian movement model for the evaluation of mixed indoor-outdoor poster campaigns. In: Proc. of the Third International Workshop on Pervasive Advertising and Shopping. (2010)
6. Li, M., Konomi, S., Sezaki, K.: Understanding and modeling pedestrian mobility of train-station scenarios. In Sabharwal, A., Karrer, R., Zhong, L., eds.: WINTECH, ACM (2008) 95–96
7. Pels, M., Barhorst, J., Michels, M., Hobo, R., Barendse, J.: Tracking people using Bluetooth. Implications of enabling Bluetooth discoverable mode. Technical report, University of Amsterdam (2005)
8. Hallberg, J., Nilsson, M., Synnes, K.: Positioning with Bluetooth. In: 10th International Conference on Telecommunications. Volume 2. (2003) 954–958
9. Andrienko, N., Andrienko, G., Stange, H., Liebig, T., Hecker, D.: Visual analytics for understanding spatial situations from episodic movement data. KI - Künstliche Intelligenz (2012) 241–251
10. Utsch, P., Liebig, T.: Monitoring Microscopic Pedestrian Mobility Using Bluetooth. In: Proceedings of the 8th International Conference on Intelligient Environments, IEEE Press (2012) 173–177
11. Hagemann, W., Weinzerl, J.: Automatische Erfassung von Umsteigern per Bluetooth-Technologie. In: Nahverkerspraxis. Springer-Verlag Berlin Heidelberg, Berlin Heidelberg (2008)
12. Leitinger, S., Gröchenig, S., Pavelka, S., Wimmer, M.: Erfassung von Personenstrmen mit der Bluetooth-Tracking-Technologie. In: Angewandte Geoinformatik 2010. 15 edn. Addison Wesley Longman Inc., New York (2010)
13. Stange, H., Liebig, T., Hecker, D., Andrienko, G., Andrienko, N.: Analytical Workflow of Monitoring Human Mobility in Big Event Settings using Bluetooth. In: Proceedings of the 3rd International Workshop on Indoor Spatial Awareness, ACM (2011) 51–58
14. Sammon, J.W.: A nonlinear mapping for data structure analysis. IEEE Transaction on Computers **18**(5) (May 1969) 401–409
15. Gong, X., Wang, F.: Three improvements on knn-npr for traffic flow forecasting. In: Proceedings of the 5th International Conference on Intelligent Transportation Systems, IEEE Press (2002) 736–740
16. May, M., Hecker, D., Körner, C., Scheider, S., Schulz, D.: A vector-geometry based spatial knn-algorithm for traffic frequency predictions. In: Data Mining Workshops, International Conference on Data Mining, Los Alamitos, CA, USA, IEEE Computer Society (2008) 442–447
17. Neumann, M., Kersting, K., Xu, Z., Schulz, D.: Stacked gaussian process learning. In: Proceeding of the 9th IEEE International Conference on Data Mining (ICDM 2009), IEEE Computer Society (2009) 387–396
18. Liebig, T., Xu, Z., May, M., Wrobel, S.: Pedestrian quantity estimation with trajectory patterns. In: Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases ECML PKDD 2012, Part II, LNCS 7524, Springer (2012) 629–643
19. De Raedt, L.: Logical and Relational Learning. Springer (2008)
20. Getoor, L., Taskar, B., eds.: Introduction to Statistical Relational Learning. The MIT Press (2007)
21. Yu, K., Chu, W., Yu, S., Tresp, V., Xu, Z.: Stochastic relational models for discriminative link prediction. In: Neural Information Processing Systems. (2006)
22. Chu, W., Sindhwani, V., Ghahramani, Z., Keerthi, S.: Relational learning with gaussian processes. In: Neural Information Processing Systems. (2006)

23. Rasmussen, C.E., Williams, C.K.I.: Gaussian Processes for Machine Learning. The MIT Press (2006)
24. Kondor, R.I., Lafferty, J.D.: Diffusion kernels on graphs and other discrete input spaces. In: Proceeding of the International Conference on Machine Learning. (2002) 315–322
25. Krause, A., Guestrin, C., Gupta, A., Kleinberg, J.: Near-optimal sensor placements: maximizing information while minimizing communication cost. In: Proceedings of the 5th international conference on Information processing in sensor networks. IPSN '06, New York, NY, USA, ACM (2006) 2–10
26. National Institute of Standards and Technology: Secure Hash Standard. National Institute of Standards and Technology, Washington (2002) Federal Information Processing Standard 180-2.
27. Rigoutsos, I., Floratos, A.: Combinatorial pattern discovery in biological sequences: The teiresias algorithm. Bioinformatics **14**(1) (1998) 55–67
28. Kisilevich, S., Keim, D., Rokach, L.: A novel approach to mining travel sequences using collections of geotagged photos. In Painho, M., Santos, M.Y., Pundt, H., Cartwright, W., Gartner, G., Meng, L., Peterson, M.P., eds.: Geospatial Thinking. Lecture Notes in Geoinformation and Cartography. Springer Berlin Heidelberg (2010) 163–182
29. Liebig, T., Körner, C., May, M.: Fast visual trajectory analysis using spatial bayesian networks. In: ICDM Workshops, IEEE Computer Society (2009) 668–673
30. Liebig, T., Körner, C., May, M.: Scalable sparse bayesian network learning for spatial applications. In: ICDM Workshops, IEEE Computer Society (2008) 420–425
31. Naini, F.M., Dousse, O., Thiran, P., Vetterli, M.: Population size estimation using a few individuals as agents. In: Proceedings of the International Symposium on Information Theory, IEEE (2011) 2499–2503