A General Pedestrian Movement Model for the Evaluation of Mixed Indoor-Outdoor Poster Campaigns

Thomas Liebig*, Hendrik Stange*, Dirk Hecker*,
Michael May*, Christine Körner*, and Urs Hofmann‡

*Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS,
Schloss Birlinghoven, 53757 Sankt Augustin, Germany
`{thomas.liebig,hendrik.stange,dirk.hecker,`
`michael.may,christine.koerner}@iais.fraunhofer.de`
‡Swiss Poster Research Plus AG, Höschgasse 70, CH-8008 Zürich, Switzerland
`u.hofmann@spr-plus.ch`

Abstract. Over the last few years new measurement technology has revolutionized the performance measurement in outdoor advertising. A handful of pioneer countries trace personal mobility now via GPS devices, which allows for precise performance results of arbitrarily positioned outdoor poster campaigns. However, GPS technology has the drawback that it cannot be applied indoors due to signal loss. In Switzerland and Germany many valuable posters are situated in public buildings such as train stations or shopping malls and their evaluation is of high interest. In this paper we therefore present a new approach for the evaluation of mixed indoor-outdoor campaigns. Our approach consists of a general pedestrian movement model in restricted spaces which can be integrated into standard trajectory evaluation. Our approach has been implemented for 27 major train stations in Switzerland.

1 Introduction

Outdoor advertisement is one of the oldest advertising media and plays an important role in the advertisement industry. In 2008 the turnover was 684 million CHF (about 460 million Euro) in Switzerland and 805 million Euro in Germany [1, 2].

In recent years the market has changed rapidly. The change is predominately caused by two factors, namely the competition with other advertising media and the emergence of digital media. Outdoor advertisement competes with other media including television, radio and press, and also more modern ones as online ads and direct mailing. To become incorporated by media planners in an advertisement mix, transparent measures are needed for the performance of a campaign. Typical measures are (1) the *coverage* or *reach* of a campaign; this is the percentage of persons within a target group defined by socio-demographic attributes that has had contact with a campaign in a certain time interval (often one week). And (2) the *number of contacts* this group has had.

Improved methods for audience measurement have become available in the last years due to technological advances and improved methodology. E.g., for performance measurement of cars and pedestrians on the street, GPS technology has established itself as a new standard in Switzerland and Germany, greatly improving the possibilities of fine-grained media planning [3, 4]. Other countries are currently preparing GPS studies, and it can be expected to become a world-wide standard. The second cause for a rapidly changing market is the emergence and increasing importance of digital media, which is the fastest growing segment in this market. Digital billboards and signs can deliver both static and animated content. Content can be changed dynamically during the day, allowing to fit the advertisement to varying target audiences that are present at different time periods. Actually, these two factors are related, since the growing importance of digital media and pervasive advertising has to be accompanied by even more fine-grained methods for performance evaluation. In the future, it will be necessary not just to calculate average performance per week, but to take account of the day of the week and even time of the day, to allow for customized advertising.

In this paper, we will focus on a problem that has not been sufficiently addressed by previous methods and that can be extended to dynamic performance measurement in the future. This is performance measurements for billboards that are placed indoors, in buildings. Performance of these indoor posters is highly interesting for media planners because, e.g. in Switzerland, about 2,600 of all posters are located in train stations. These locations are considered among the most valuable ones. We will investigate the special case of *inner train station campaigns* in Switzerland in this paper. We are going to estimate poster reach for campaigns within the biggest 27 Swiss train stations. The challenge in this case results from the fact that due to signal loss caused by the building, the GPS trajectories (as used for streets) just describe which persons enter the building; inside the building itself valid GPS positions are rare and generally not available. Thus, we do not know which person has contact with a particular poster or indoor campaign. It is important that the performance values for train stations are *embeddable* in the GPS model. Additionally, in a future world of pervasive advertising we need more complete information at an even finer spatial resolution about the trajectories and movement patterns of persons in buildings.

Any study of performance measures is in need of empirical data. In our case tracking pedestrians with cameras seems to be a perfect solution at first sight. But in practice it turns out that the use of video technology is often not permitted in train stations due to privacy restrictions. Radio Frequency Identification (RFID) and Bluetooth [5] technology may also be used for tracking, but this becomes expensive and requires an additional infrastructure for deployment of the necessary hardware. A third option for trajectory recording are interviews. However, they are very time consuming and thus expensive.

Our approach is based on (1) obtaining a number of relatively inexpensive frequency counts manually, and (2) to generate a general model for indoor pedestrian movements based on frequency counts and a network of the possible pathways through a train station. In the next step (3) we infer reach values from this model by combining it with the GPS measurements available from outdoor measurements. In this paper, the main focus is on step (2). The paper is structured as follows. In section 2 we discuss related work. Next we describe in section 3 how empirical data has been collected. Section 4 is the main part of the paper and describes the model for pedestrian flow estimation. Section 5 discusses some validation issues and section 6 addresses reach estimation. We conclude with a section on future work.

2 Related Work

To the best of our knowledge, previous work concerning poster campaign evaluation only deals with outdoor campaigns. For the estimation of traffic flows in general a number of methods and algorithms exists in the literature. Initially being an operation research problem concerned with logistics and transportation issues, with increasing map-sizes it became a problem for spatial data mining. First, there exists a large group of probabilistic microsimulation models including Monte Carlo methods [6], Markov models [7], cellular automata [8] and multi agent simulation [9]. Second, there are large-scale macroscopic algorithms for frequency prediction in extensive road networks [10]. Finally, there is a group of regression algorithms [11–13].

Each of these models and algorithms makes certain assumptions on the traffic behaviour, reflecting different aspects of real-world traffic. Microsimulation models are useful for evacuation planning and obstacle detection. The regression models quoted above tend to represent real-world traffic more accurately, because they can be easily calibrated with measurement data. Moreover, the Kirchhoff laws are automatically fulfilled by any solution at each crossing. This means that the solution is always valid, because the number of incoming equals the number of outgoing people, and the model fits the measurement data. Nevertheless, they are not easy to apply to large traffic networks with few empirical data. For one reason, the computation time of the method increases rapidly while the accuracy becomes worse.

However, since train stations and buildings are small, compared to cities or highway road networks, our approach is based on a regression model. This ensures that the resulting model conforms to measurements and represents a valid pedestrian flow. The known regression approaches [11–13] focus on frequency estimation, but not on complete trajectory reconstruction. The method we present in the next two sections achieves both at once and results in a general pedestrian movement model based on a small sample of empirical frequency counts.

3 Empirical Data Collection

Some of the pedestrian models presented in section 2 require detailed representations of the accessible space. Applications for such models are emergency and evacuation planning, capacity analysis or obstacle detection. As we are only interested in frequencies and information whether persons pass certain locations, we do not require such a detailed model. Therefore, a graph representation of the floors, stairways and junctions contains enough information for our task. For our project, we built such graphs for all of the considered 27 Swiss train stations. Every junction is represented by a vertex and the connecting floors are represented by edges. Trajectories through the station may then be described by a sequence of edges, starting and ending at an entrance or platform.

After doing a pre-study, we concluded that counting the number of people manually at several positions (using a smart phone application for data entry) is the most cost-effective method for data collection. As noted in the introduction, using video cameras was not feasible because of privacy constraints. To decrease the influence of the day of week on the measurements, we repeated the measurements at 3 different days. As the number of "sensors" is limited, we had to select locations for counting in advance using the traffic network of the train stations. Therefore, we located sensors at the most important junctions and stairways. Figure 1 depicts the measured edges at Zurich central station.

To assist manual counting and to simplify post-processing of measurements, we developed a smart phone application (figure 1) which records clicks of the surveying person - each click represents the number of pedestrians passing in a specified direction - along with its timestamp. This enables an easy storage of the data in a database. Thus, we know how many people passed at which time into which direction. In an early prototype, we encountered the problem of mixed directions; therefore we added visual hints to the smartphone application as well as to the map. To distinguish directions, the colours red and green are used in our application.

To be able to compare the empirical raw data of pedestrians at measurement locations, e.g. an average number of pedestrians for a complete week, post-processing is necessary: after merging the measurements, frequencies are weighted and aggregated according to the time interval and day they were taken. As a result, every measured location in the train station has associated with it a number of pedestrians that may be compared against any other location. This is important for ranking locations or tracking segments within the building, which is a first feature of our general mobility model.

4 Pedestrian Flow Estimation

For segments where empirical measurements have been taken, the frequencies are known. Our task is now to estimate frequencies for the unobserved segments, and to build a general pedestrian indoor movement model that is useful for poster and campaign evaluation.

In contrast to other regression models [11–13] that do not give trajectories but just frequencies, we tackle both questions at the same time, using a two stage regression approach. In a first step, we enumerate all plausible routes through the building and collect them in a *route set*. For example, at the main station in Zurich, there are about 380,000 conceivable routes. Non-plausible routes are eliminated, among them circular routes. Afterwards, we assign frequencies to

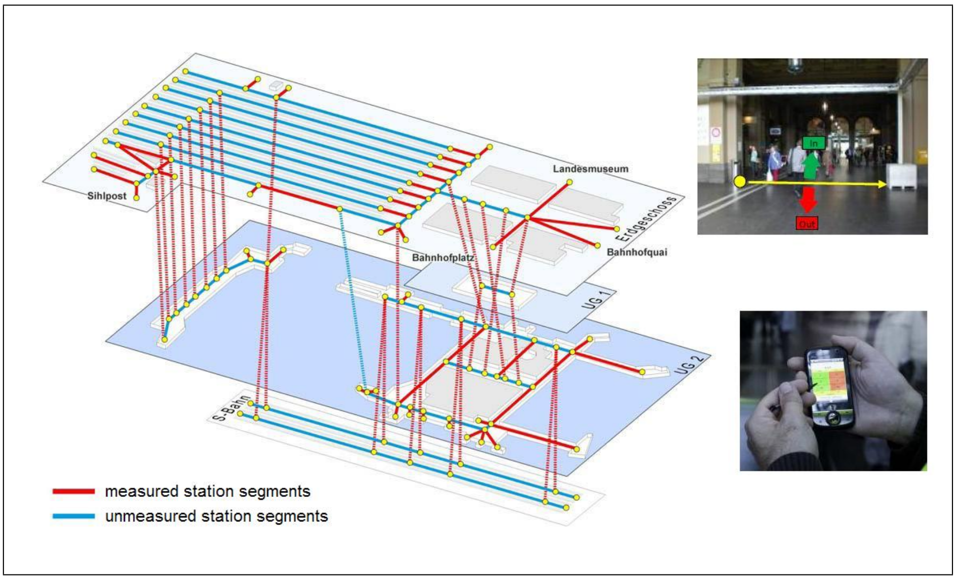

Fig. 1. Selected Locations for Frequency Measurements

each route, based on the measurements. The measurements serve as frequency targets in this process. The purpose of this assignment procedure is to find the optimal combination of routes that fulfills all frequency targets. Simplified, our approached may be seen as follows: from all possible routes take those that fit best, add this to the, initially empty, solution set and repeat this process until the target frequencies are met as accurately as possible. In our software implementation, we represent the set of all plausible routes by a binary matrix. It encodes which edges are contained in each route. Every column represents an edge and every row a route. Except for the cases where the $edge_i$ is contained in the $route_j$ all elements of the matrix ($A = (a_{ji})$) are zero. Because routes containing circles are eliminated as non-plausible beforehand, every edge occurs at most once within a route; hence, a binary representation is adequate. Using this matrix, the target frequencies F_{edges} can be rewritten as sum of the passing route frequencies, as shown in the following equations:

$$A = (a_{ji})^{0 \leq i \leq |edges|}_{0 \leq j \leq |routes|}$$

$$a_{ji} = \begin{cases} 1 & \text{if } edge_i \in route_j, \\ 0 & \text{otherwise} \end{cases}$$

$$F_{edge_i} = \sum_{1 \leq j \leq |routes|} a_{ji} * F_{route_j}$$

$$F_{edges} = A^T \times F_{routes}$$

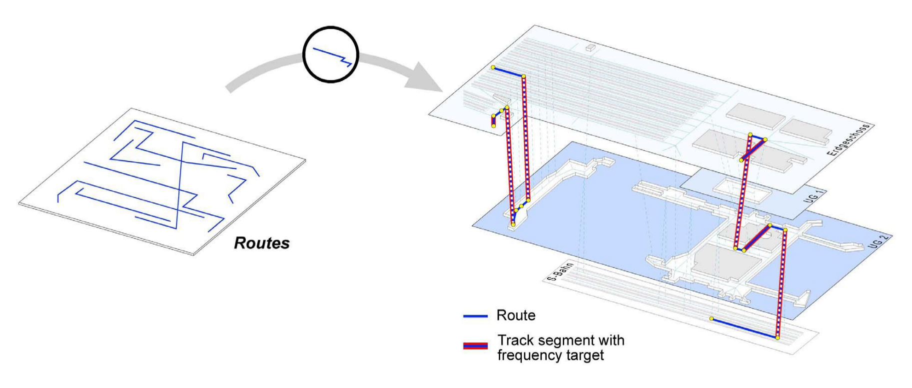

Fig. 2. Assignment of frequencies

To find the best set of route frequencies, we use *least square regression*. Thus, the assignment problem becomes a convex optimization program described by the following equation.

$$
\begin{aligned}
&F^*_{routes} = \arg\min \left|\left| A^T \times F_{routes} - F_{edges_{target}} \right|\right| \\
&F^*_{routes} \geq 0
\end{aligned}
$$

We are looking for the combination of route frequencies F^*_{routes} that involves the smallest difference to the target values. It might be possible that there does not exist a single solution but a solution set, containing multiple of these route frequencies. Taking empirical observations into account, we tend to that solution which uses the most distinct paths. This ensures the resulting model to cover possible route candidates with an appropriate low instead of zero frequency. Furthermore, we give preference to faster walk-throughs through the train station, as detours are unusual for pedestrian movements [9]. We solve this convex program using a gradient-descent algorithm.

As a result we obtain for every modeled train station (1) a set of routes crossing that station and (2) the number of people walking on each route. Figure 3 gives an example for Zurich central train station. With this information we are able to calculate edge frequencies F_{edge_i} for any $edge_i$ in the station by summing over the route frequencies F^*_{routes}, no matter whether the edge has been measured empirically or not. This yields the general pedestrian movement model based on empirical measurements we aim for. It enables us to denote frequencies at any edge and gives trajectories also at unobserved segments.

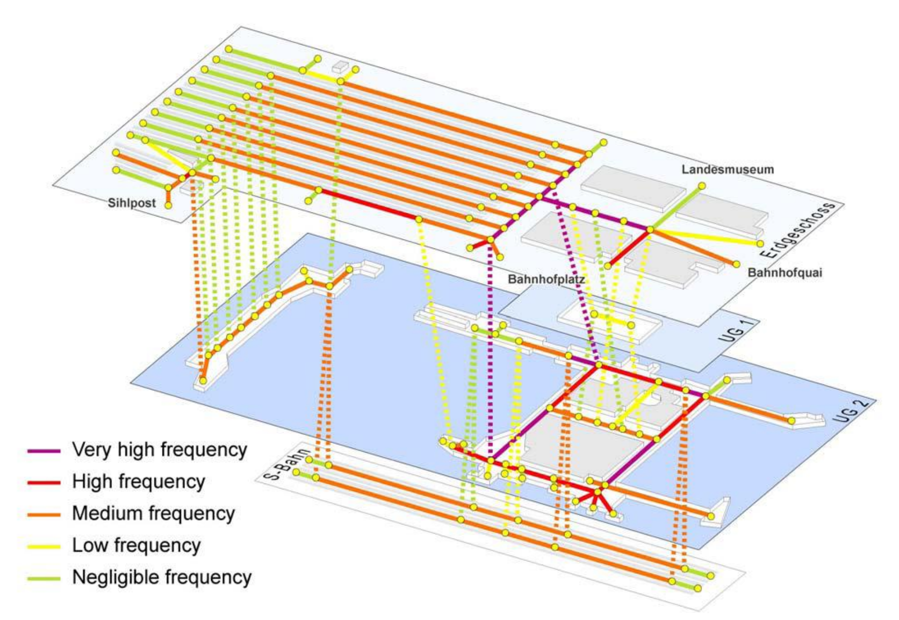

Fig. 3. Estimated Frequencies in Zurich

5 Robustness Against Outliers

The empirical measurements are not all taken in the same time interval. Thus, the combination of these frequencies in target frequencies $F_{edges_{taget}}$ must lead to contradictions (if time is ignored). At junctions, they can be easily recognised in the raw data by violations of Kirchhoff's law, whereby the number of incoming people has to equal the number of the outgoing ones. In general, this problem may also arise along multiple junctions. In this case, it is harder to pre-identify the contradictions. Therefore, we require the frequency estimation algorithm to recognise such cases automatically and to eliminate them in the model, if required.

Our approach, described in section 4, fulfills all of these criteria. The Kirchhoff law holds automatically, because the enumerated route set does: Every single route holds the constraint at any junction that the number of incoming people equals the number of outgoing. Multiplying with the route frequencies, our algorithm increases the frequency of a route, but nevertheless the equilibrium remains fulfilled. If each single route fulfills this constraint, the set of all routes including their final frequencies also does. Therefore, in our general pedestrian model Kirchhoff's law holds. Furthermore, small perturbances in the frequencies are corrected by the least squares regression. Without the need for a pre-analysis of outliers the resulting frequencies are chosen such that the differences at the measurement locations are minimal. This is defined by the objective function of the program in section 4.

6 Performance Measures

In order to apply our pedestrian movement model for the evaluation of mixed indoor-outdoor poster campaigns, we still need to integrate it with the GPS mobility data. This means we have to assign for each GPS person who enters a railway station a corresponding route through the station. To achieve this goal, we define three subsequent steps: (1) visit identification, (2) route assignment and (3) performance evaluation.

In the first step we identify all test persons within the GPS sample visiting a train station. Based on GPS trajectories of over 10,000 test persons recorded over a period of one week per person, we isolate all tracks in the vicinity of a train station using buffers and the spatial join operation "intersect". As GPS signals may be noisy, we apply an individually sized buffer to each of the train station geometries, reflecting its specific local setting. The resulting candidate set, however, contains not only potential rail travelers but also regular pedestrians, car drivers and passengers passing by the station without entering it. We therefore apply a complex multi-level filtering process which identifies the visitors of a train station using, for instance, speed curves, the course of movement and time spent inside the geometric extension of the train station. Knowing all visits to a railway station completes step 1.

Step 2 is the assignment of each visit to one of the routes underlying the pedestrian movement model. The challenge of this task is to find an optimal distribution of personalized routes given the route frequencies F_{routes}. We do this iteratively by drawing routes from the route set and considering the projected weight and socio-demographic information of each test person being assigned to that particular route. At the end of this process each GPS trajectory containing a visit to a train station as identified in the previous step, has been assigned a route through the corresponding train station.

Finally in step 3 we weight poster contacts and calculate performance measures of mixed in- and outdoor campaigns. Similar to the performance evaluation of outdoor posters, we consider individual visibility criteria at each poster site. Routes passing the visibility area are weighted according to the contact quality, depending e.g. on the viewing direction or clustering of panels. Given weighted contacts for each indoor poster and visiting person, we can estimate total contacts and reach of a mixed indoor-outdoor campaign using the same algorithmic background as for outdoor campaigns. The selection of a campaign and of a target audience determines all relevant (indoor and outdoor) poster contacts and the application of Kaplan-Meier compensates for missing measurement days in the GPS data as described in [4].

7 Conclusion and Future Work

In this paper we developed a method that allows performance measurements for billboards that are placed indoors. We focussed on 2,600 poster sites in railway stations as those are being seen as one of the most valuable over all. The challenge results from GPS signal loss inside buildings. Our approach includes the development of a general pedestrian model based on empirical data. This mobility information has been integrated with existing GPS mobility data, allowing to infer reach values and weighted contacts. We applied our approach to 27 major Swiss train stations.

Although we showed how to implement a general movement model within the train station which is used for poster campaign evaluation, we do not model time, so far. Our indoor model is as static as the cited outdoor evaluation methods. While for the current needs of outdoor performance measurements a static model is sufficient, in a future world of pervasive advertisement a dynamic model that uses the time dependencies of the measurements, would bring immense benefits. In the future, modelling them with Gaussian Processes enriched with relational data sources, as train schedules or text messages like newstickers, is a promising method. It will enhance our method and allow for more target group specific campaign measurements in the future. In combination with persistent frequency sensors, we aim to model real-time poster evaluations.

References

1. Stiftung Werbestatistik Schweiz: Werbeaufwand Schweiz, Erhebungsjahr 2008 (Advertising expenditure Switzerland, survey year 2008) (2009)
2. Fachverband Außenwerbung e.V.: Netto-Werbeeinnahmen erfassbarer Werbeträger in Deutschland, 2000-2008 (Net turnover of confirmable advertising media in Gemany, 2000-2008) (2009)
3. May, M., Körner, C., Hecker, D., Pasquier, M., Hofmann, U., Mende, F.: Handling missing values in gps surveys using survival analysis: a gps case study of outdoor advertising. In: ADKDD '09: Proceedings of the Third International Workshop on Data Mining and Audience Intelligence for Advertising, New York, NY, USA, ACM (2009) 78–84
4. Pasquier, M., Hofmann, U., Mende, F.H., May, M., Hecker, D., Körner, C.: Modelling and prospects of the audience measurement for outdoor advertising based on data collection using gps devices (electronic passive measurement system). In: Proceedings of the 8th International Conference on Survey Methods in Transport. (2008)
5. Alt, F., Balz, M., Kristes, S., Shirazi, A.S., Mennenöh, J., Schmidt, A., Schröder, H., Goedicke, M.: Adaptive user profiles in pervasive advertising environments. In: Proceedings of the 3rd European Conference on Ambient Intelligence (AmI'09). Volume 5859 of Lecture Notes in Computer Science., Springer (2009) 276–286
6. Borgers, A., Timmermans, H.J.P.: City centre entry points, store location patterns and pedestrian route choice behaviour: A microlevel simulation model. Socio-Economic Planning Sciences **20**(1) (1986) 25–31
7. Yu, G., Hu, J., Zhang, C., Zhuang, L., Song, J.: Short-term traffic flow forecasting based on markov chain model. In: Proceedings of the IEEE Intelligent Vehicles Symposium, 2003., IEEE (2003) 208–212
8. Blue, V.J., Adler, J.L.: Cellular automata microsimulation for modeling bi-directional pedestrian walkways. Transportation Research Part B: Methodological **35**(3) (March 2001) 293–312
9. Rindsfüser, G.: Simulation des Aktivitätenplanungsprozesses mit Methoden der Künstlichen Intelligenz. Straßenverkehrstechnik **49**(3) (2005) 145–153

10. May, M., Hecker, D., Körner, C., Scheider, S., Schulz, D.: A vector-geometry based spatial knn-algorithm for traffic frequency predictions. Data Mining Workshops, International Conference on Data Mining **0** (2008) 442–447
11. Pushkarev, B., Zupan, J.M.: Pedestrian travel demand. Highway Research Record **355**(HS-011 999) (1971) 37–53
12. Tanyimboh, T.T., Templeman, A.B.: Calculating maximum entropy flows in networks. Journal on Operation Research **44**(4) (1993) 383–396
13. Yassin-Kassab, A., Templeman, A.B., Tanyimboh, T.T.: Calculating maximum entropy flows in multi-source, multi-demand networks. Journal on Engineering and Optimization **31**(6) (1999) 695–729