

Crowd Replication: Sensing-Assisted Quantification of Human Behaviour in Public Spaces

Hemminki, Samuli
University of Helsinki

Kuribayashi, Keisuke
University of Tokyo

Konomi, Shin'ichi
Kyushu University

Nurmi, Petteri
University of Helsinki

他

<https://hdl.handle.net/2324/4755239>

出版情報 : ACM Transactions on Spatial Algorithms and Systems (TSAS). 5 (3), pp.1-34, 2019-09.
Association for Computing Machinery: ACM
バージョン :
権利関係 : © 2019 Association for Computing Machinery.



Crowd Replication: Sensing-Assisted Quantification of Human Behaviour in Public Spaces

SAMULI HEMMINKI, University of Helsinki
 KEISUKE KURIBAYASHI, University of Tokyo
 SHIN'ICHI KONOMI, Kyushu University
 PETTERI NURMI, University of Helsinki and Lancaster University
 SASU TARKOMA, University of Helsinki

A central challenge for public space design is to evaluate whether a given space promotes different types of activities. In this paper, as our first contribution, we develop crowd replication as a novel *sensor-assisted* method for *quantifying human behaviour* within public spaces. In crowd replication, a researcher is tasked with recording the behaviour of people using a space while being instrumented with a mobile device that captures a sensor trace of the replicated movements and activities. Through mathematical modelling, behavioural indicators extracted from the replicated trajectories can be extrapolated to represent a larger target population. As our second contribution, we develop a novel highly accurate pedestrian sensing solution for reconstructing movement trajectories from sensor traces captured during the replication process. Our key insight is to tailor sensing to characteristics of the researcher performing replication, which allows reconstruction to operate robustly against variations in pace and other walking characteristics. We validate crowd replication through a case study carried out within a representative example of a metropolitan-scale public space. Our results show that crowd-replicated data closely mirrors human dynamics in public spaces, and reduces overall data collection effort while producing high quality indicators about behaviours and activities of people within the space. We also validate our pedestrian modelling approach through extensive benchmarks, demonstrating our approach can reconstruct movement trajectories with high accuracy and robustness (median error below 1%). Finally, we demonstrate that our contributions enable capturing detailed indicators of liveliness, extent of social interaction, and other factors indicative of public space quality.

Additional Key Words and Phrases: Urban computing, crowd replication, mobile sensing, public spaces, evaluation, pedestrian modeling, motion sensing

ACM Reference format:

Samuli Hemminki, Keisuke Kuribayashi, Shin'ichi Konomi, Petteri Nurmi, and Sasu Tarkoma. 1. Crowd Replication: Sensing-Assisted Quantification of Human Behaviour in Public Spaces. *ACM Trans. Spatial Algorithms Syst.* 1, 1, Article 1 (January 1), 34 pages.
 DOI: 0000001.0000001

1 INTRODUCTION

Public spaces serve an important role in contemporary cities. Among others, they provide a medium for passive and active engagement, and a place for discovery and reflection [11]. Public spaces have

The work was partially supported by Nokia foundation scholarship, by the Finnish Academy of Science under the projects Sampling in Pervasive Sensing Systems 296139, and Context Sensing for Security 303825, by JST CREST Grant Number JPJPMJCR1411, and by JSPS KAKENHI Grant Number JP17K00117. We thank the members of Spatial Planning and Design Laboratory at the University of Tokyo for sharing the pedestrian count data. The work only reflects the authors' views.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 1 ACM. 2374-0353/1/1-ART1 \$15.00

DOI: 0000001.0000001

also been linked with quality of life indicators [12]. Indeed, well-designed public spaces can promote individual well-being, and contribute to social and environmental values of neighbourhoods. With ever increasing motorization and urbanization, the role of public spaces has been transforming and their role as hubs for human activities has diminished. To counter this trend, urban planners, activists and architects have been advocating for new ways of enriching human experiences in public spaces by designing the environments so that positive qualities and liveliness of public spaces are increased [58].

Public space quality is intrinsically linked with the way it is used. According to Gehl [17], uses of public space can be categorized into necessary, optional, and social activities, with a successful space being one that actively promotes optional and social activities. Currently, however, urban planners struggle with quantifying whether a particular space promotes these activities. This is because current techniques for public space assessment are either too labour-intensive and provide insufficiently detailed information about usage of space, or require costly investment and high level of technical expertise. Indeed, despite technological advancements, direct observation remains the dominant approach for quantifying public space usage [26, 37, 57]. In direct observation, the researcher manually observes and logs activities of people within the space. From these observations, indicators related, e.g., to the amount of mobility and activities within the space can be derived [37]. The main limitations of this approach are the limited granularity of the collected information, and labour-intensiveness of the data collection process, usually resulting in a small sample size. While a range of pervasive sensing technologies, e.g., visual surveillance [14, 40], aerial photography [18], location tracking [41, 52], crowd sensing [46, 53, 56, 59], and analysis of cellular records [8], can support data collection, they have not been widely adopted due to high resource cost, need for expertise in interpreting the data, and privacy concerns. Additionally, pervasive sensing techniques are vulnerable to biases that limit generalizability of the conclusions that can be drawn. For example, visual surveillance is prone to losing data due to occlusions or illumination conditions whereas crowd sensing captures only users carrying devices with an active WiFi or Bluetooth interface. Accordingly, data collected through pervasive sensing is prone to contain temporal or spatial gaps, and may be biased to a specific sub-population of the people using a space. To address the limitations of existing approaches, *an approach that captures unbiased and detailed information of human behaviour with minimal resource cost is required.*

In this paper, as our first contribution, we develop *crowd replication* as a novel *sensor-assisted* method for quantifying human behaviours within a public space. The idea in crowd replication is to augment direct observation through mobile sensing, enabling capture of detailed information about human behaviour. In crowd replication, the *researcher* carrying out an investigation is engaged into replicating the behaviours of target subjects while carrying a mobile device that collects a sensor trace of the movements and activities of the subjects; see Figure 2b for an illustration.

From the sensor trace, we extract detailed movement trajectories and motions of people. These are then combined with manual annotations to derive indicators that relate motion characteristics (e.g., walking pace, motion intensity, and heading) with public space use on fine-grained spatial and temporal resolution. To relate these characteristics with the entire target population, the derived indicators can be extrapolated to capture the aggregate behaviour of the entire target population. Compared to pervasive sensing, crowd replication provides better control over sample selection and requires little additional investments or technical expertise. Indeed, a key design goal for crowd replication has been to make it easily adoptable by researchers and practitioners versed in direct observation. Another benefit of crowd replication is that it can be carried out using off-the-shelf smartphones without installing any sensors or other hardware within the public space being evaluated. This contrasts with mobile crowdsourcing where users need to download and install

an application on their smartphones. As crowd replication collects measurements only from the researcher(s) conducting the evaluation, privacy of people using the space can be protected. At the same time, this helps to improve quality of sensor data as placement of mobile device collecting sensor data can be better controlled. The main alternative to crowd replication, visual surveillance, provides a higher amount of measurements but requires either expensive visual analytics software or large amounts of manual effort in annotating video data, making it costly and labour-intensive. Visual surveillance also suffers from high deployment cost due to privacy considerations, the need to obtain necessary permits, and the need to find unobstructed locations for placing the cameras. We stress, however, that crowd replication does not aim at removing the need for visual surveillance, pervasive sensing, or other techniques. Rather, crowd replication has been designed as a cost-effective and easy-to-adopt technique that researchers can use to capture high-quality data to quantify behaviour within public spaces. Crowd replication can also be used to complement other techniques as it provides detailed and unbiased information on the behaviour of people within the target site. Such samples could then be leveraged to interpret, validate and correct data and observations made through other data collection.

A fundamental challenge for crowd replication is to ensure that the collected data is sufficiently accurate and representative of the behaviour of people using the space. To ensure crowd replication captures high quality information of people using public space, as our second contribution, we develop a novel highly accurate pedestrian sensing solution for reconstructing movement trajectories from sensor traces captured during the replication process. What makes this particularly challenging is the high variability of motion, with walking pace, heading changes, step intensity and other walking characteristics differing depending on the activities of the people being monitored. Capturing these subtle variations in walking characteristics is essential for understanding uses of public space as these changes are highly correlated with different types of activities taking place within the space [35]. Further complications are caused by stationary periods which can be frequently interleaved with periods of motion. To address accuracy requirements of crowd replication, our pedestrian modelling approach learns a *personalized* model of the stride patterns of a user. In crowd replication, only a small set of users are responsible for collecting data, i.e., the researchers carrying out the replication. Furthermore, the devices collecting sensor measurements are primarily used for sensor data collection instead of serving as personal communications devices. This provides better control over device placement and orientation, and reduces the impact of spontaneous user interactions and noise on the sensor measurements. Our sensing solution takes advantage of these factors, mining the collected sensor measurements for step cycles which then can be associated with indicators related to motion intensity, stride length, and walking pace. This contrasts with existing research on pedestrian behaviour sensing which has predominantly targeted scenarios where the focus is on generalization over various users and unknown device use and placement at the expense of sensing accuracy [7, 22].

We validate and demonstrate the benefits of crowd replication through a case study carried out within a representative example of a metropolitan scale public space. We first demonstrate that crowd replication captures data that is representative of the target population and their behaviours. We show this by assessing the convergence of the distributions of demographic variables and motion trajectories captured through crowd replication. Our results show that for most variables of interest (demographics and trajectories) the distributions can be learned with a reasonably small sample (50 – 100 replicated users). As can be expected, the slowest convergence is witnessed with infrequently occurring activities (social activities in our experiments), which can require up to 200 replicated users. This amount of data can be reasonably gathered in a matter of hours. For example, in our case each researcher collected 25 – 30 participants per hour, which means that

two researchers could collect sufficient amount of data in around 4 hours. In contrast, manually annotating 30 minutes of video surveillance data required over 50 hours of work, and only produced aggregate level information on the behaviour of people using the space. As the second step of our evaluation, we evaluate the accuracy of the collected data by comparing dwell indexes extracted from crowd replicated data against those extracted from manually annotated video data [21]. Our results show a high level of correspondence between the two sets of indicators (0.914 correlation). We also separately assess the accuracy of our pedestrian modelling solution, demonstrating that it provides better accuracy than current state-of-the-art techniques. The results demonstrate that our approach is particularly effective against variations in walking pace, providing highly accurate and robust performance while the performance of state-of-the-art solutions degrades. As variations in pace often relate to different activities, e.g., people walking in groups tend to have slower walking pace, our approach is thus better capable of capturing subtle nuances in the walking patterns of people using the space – a critical requirement for assessing optional and social activities. As the third part of our evaluation, we compare data quality and overall effort of crowd replication against video tracking and pervasive sensing techniques. Our results demonstrate that crowd replication can significantly increase sample sizes compared to direct observation, while at the same time being more cost effective than current urban monitoring solutions. As the final step of evaluation, we demonstrate that our sensing approach enables crowd replication to capture spatiotemporally fine-grained indicators about human behaviour within public spaces. Specifically, we demonstrate that reliable and detailed indicators of liveliness and extent of social interaction can be obtained with the proposed method. The resulting data from our experiments, together with details about the spatial area and model parameters has been made freely available for research purposes at http://akiyatoolbox.org/crowd_replication/.

Summary of Contributions

The contributions of the paper are summarized as follows:

- We develop crowd replication as a novel and resource-efficient method for quantifying human behaviour within public spaces.
- We develop a novel, highly accurate pedestrian modelling solution, and demonstrate that it operates robustly against variations in walking pace.
- We validate crowd replication through a case study carried out in a representative urban space, demonstrating that it results in data that is representative of the target population and their behaviours.
- We critically compare work effort, data granularity, sample size, potential biases, and other factors between crowd replication and other data collection mechanisms. Our comparison shows that crowd replication provides a good trade-off between data collection effort and accuracy of collected data. The deployment and analysis costs of crowd replication are also moderate, making it a cost-effective alternative that can replace or supplement other techniques within up to moderately sized public spaces.
- We demonstrate that crowd replication enables capturing indicators that provide fine grained information about the movements, activities, and human experience within an urban space. Examples of such indicators include liveliness, extent of social interaction, and temporal variations in motion patterns.
- We make the data used in our experiments freely available for research purposes at http://akiyatoolbox.org/crowd_replication/.

2 RELATED WORK

Our research draws inspiration from three threads of previous research: (i) methods for evaluating public space quality; (ii) sensing assisted modeling of human behaviour; and (iii) modeling of pedestrian behaviour. In the following we describe related work in these topics.

2.1 Public Space Evaluation

To increase positive qualities of public spaces, urban planners, activists and architects have assessed key qualities of public spaces through careful direct observations [26, 37, 57]. Of particular importance in these investigations have been the different types and varieties of activities taking place, together with their distributions in terms of spatial, temporal and demographic variables. There are three widely acknowledged categories for activities of interest in public spaces: necessary, optional and social [17]. While necessary activities, such as walking to a post office, take place regardless of the quality of a public space, optional activities, such as strolling to appreciate fresh air or liveliness of streets, depend on actual and perceived qualities of the space. Social activities such as seeing and talking or playing with other people may spontaneously occur when necessary and optional activities are sustained over a period of time. Indeed, the amount and extent of activities in a public space has been shown to correlate positively with the amount of social interactions [16]. Urban planning experts have argued that successful public spaces are characterized by diverse activities by people from various age groups with the size and gender distribution of the groups also having an effect [58]. These factors help, for example, to distinguish between lively and crowded public spaces. Despite theoretical developments and several established rules of thumb, little work has proposed practical methods for uncovering spatiotemporal distributions, compositions, and dynamics of activities in public spaces. Our work addresses sensor-assisted capture of these aspects.

Traditional approaches to assessing public space quality heavily rely on direct observation and surveys. For example, researchers in architecture and urban planning have recorded pedestrian movements using surveys and direct observation to analyse movement patterns [23, 39, 49]. Mehta [36] conducted direct and walk-by observations in three neighbourhoods in Boston, manually measuring stationary and sustained activities, including social activities, at 78 block segments, each of which is approximately 15 – 18 meters in length. The collected data was analysed by deriving a liveliness index which reflects the aggregated duration of stay for all visitors in each block segment. This data collection method can quickly cause fatigue, leading into inaccuracy and missed measurements (e.g., when it is crowded) and provides limited scalability. Video cameras provide an alternative for detecting and tracking crowd activities. However, analysis of video tracking data tends to be very labour intensive and requires special expertise. Computer vision algorithms can be used to detect and track pedestrians, bicycles, and vehicles automatically [14, 40]. Nevertheless despite the expanding arsenal of video processing techniques, including recent developments on deep learning based visual analytics [55, 62], video surveillance is challenging as video resolution needs to be sufficiently high and cameras need to be placed in non-occluded locations. Additional considerations include field of view of the cameras, light conditions, high deployment costs, and privacy problems with ubiquitous video surveillance [9, 25]. Finally, installing video cameras requires permits which can be difficult to obtain due to legislation and privacy concerns. Alternatively, laser range scanners can be used to detect and track pedestrians without significant privacy invasion although it can be difficult to quantify a large public space. For example, Shao et al. [48] required eight scanners to measure a 60m × 15m public space.

2.2 Pervasive Sensing for Crowd Observation

In indoor environments, crowd monitoring can be performed using high precision indoor localization or infrastructure-based tracking. As an example, Larson [32] used RF-ID measurements to track movements of people within a retail environment. Instead of tracking people, Weppner et. al. [56] estimate overall density of people using a space from Bluetooth scans. A related approach is proposed by Schauer et. al. [46] who combine Bluetooth and WiFi scanning for crowd density estimation. These solutions only provide an estimate of the overall number of people using the space and cannot be used to characterize the spatiotemporal distribution of activities, movements, and other behaviours of the people using the space. Related commercial tracking solutions include Ubisense [52] and Quuppa HAIP [41], which provide high resolution indoor tracking, and Walkbase [53] which utilizes beacon request messages send by WiFi-enabled mobile devices to estimate crowd movements. While these technologies could be adopted for urban public spaces, they suffer from high deployment costs and require the user to carry specific hardware, or have specific settings on their mobile devices (e.g., Bluetooth/WiFi continually on). Additional concern with these technologies is that the resulting samples often contain significant and uncontrollable biases, making it difficult to draw generalized inferences from the measurements.

Mobile sensing has been shown to be useful in application domains such as transportation, environmental monitoring, and health and well-being [31]. Location tracking using GPS or cellular (GSM or CDR) data can be used for evaluating public space quality [43]. Such data can be used to analyse, e.g., macroscopic patterns of cities to estimate aggregated mobility modes, or regional events [8]. While mobile phone-based location tracking techniques are capable of capturing dynamics of people within the space, the granularity at which they can track activities is very coarse and does not adequately address the concerns of designers and planners of human-scale spaces. For example, capturing dwelling patterns or differences in activity patterns between weekdays and weekends is not feasible with these methods. Another alternative is to enhance traditional direct observation methods by using wearables, such as video cameras. As an example, architecture researchers have used portable digital cameras to track pedestrians while filming the movements of their feet [45]. The main issue with these techniques is that the analysis of the video images is extremely labour intensive and requires multiple annotators for ensuring quality of the analysis.

2.3 Pedestrian Modeling

Our approach for trajectory modeling draws inspiration from techniques for pedestrian movement modeling, and particularly microscopic agent-based modeling [3, 28, 50]. Instead of tracking participants directly, these solutions create synthetic trajectories by simulating movements of individual pedestrians. First approaches to agent-based modeling relied on cellular automata based models where agents move in a grid according to certain decision rules. In the simplest case the movements can correspond to a random walk, whereas in more realistic cases additional constraints, such as accessibility and road infrastructure information are incorporated into the model [20, 47]. Asano et al. [2] propose a variant that incorporates two types of movement rules: tactical and operational. The tactical module affects user's speed and direction of movement, whereas the operational module is responsible for micro-motions responsible for avoiding obstacles or other pedestrians. Other examples of factors that can be incorporated into the pedestrian decision model include walking speed, direction of movement, utility of different areas, and target destination of the agent; see, e.g., Hoogendoorn and Bovy [24] for an overview.

More recently, approaches that rely on generative probabilistic models have been proposed. Instead of attempting to simulate movements of individuals in a piece-meal fashion, these solutions attempt to provide a model that can be used to sample representative examples of actual trajectories.

Kitani et al. [29] and Ziebart et al. [64] use Markov decision processes (MDPs) to model pedestrian movements, which allows different actions to be given different utilities. For example, the model allows assigning higher weight to sidewalks than other areas. Other related approaches have been extensively proposed in computer vision [10, 30, 44]. A challenge with these approaches is to capture the reward structure of the participants, which requires either highly detailed trajectories or detailed information about the space where the model is applied. Also, these solutions are predominantly applicable only to visual surveillance data which, as we discuss later in the paper, has higher burden on data collection. Jitta and Klami [27] propose a model based on partial Hidden Markov Models. Instead of requiring a reward structure, their approach fits the model based on spatial features (such as dwell time or average pace length) at different parts of the space. Also the use of deep learning has been explored. As an example, Alahi et al. [1] propose a social LSTM model that integrates a so-called social layer which allows trajectory predictions to adapt to social interactions whenever multiple persons are within each other's proximity. The main limitation of both of these techniques is that they require extensive amounts of training data to produce high quality models, and that the performance of the models is sensitive to the features that are available. Also, while the models can typically capture average behaviour well, they are unable to capture deviations from this behaviour, e.g., when special events are taking place within a public space.

2.4 Summary

To summarize, while there have been advances in sensing techniques for assessing public space quality, these techniques have not been adopted by practitioners. Indeed, practitioners often have the perception that limiting data collection to sensed data leads to poorer sets of collected attributes [60] and results in inherent biases due to digital differences [63]. Our work addresses this discrepancy, providing a mechanism where researchers have control over sample selection, but at the same time can benefit from finer granularity data provided by sensing technologies.

3 SENSOR-ASSISTED QUANTIFICATION OF HUMAN BEHAVIOUR

The focus of our research is on developing a resource-efficient method for collecting high-quality and unbiased data about human behaviour within public spaces. To accomplish this, we have developed crowd replication as a novel approach for public space evaluation. In crowd replication, the researcher responsible for evaluating the public space is tasked with replicating the movements and activities of people using the space while being equipped with a smartphone that captures a sensor trace of those movements and activities. The sensor data is then combined with visual observations, e.g., about user demographics and labels describing activities performed by users. This combination of direct observation and automated sensing helps crowd replication to capture the nature and characteristics of optional and social activities [17], two of the key dimensions for assessing public space quality. In the following we first introduce the overall idea of crowd replication, after which we describe each main phase of the overall process. Finally, we detail technical aspects of the proposed pedestrian sensing technique that we have developed to support crowd replication.

3.1 Overview of Crowd Replication

Crowd replication combines mobile sensing and direct observation to provide detailed spatiotemporal information about human behaviour within a public space. The central idea in crowd replication is to task the researcher to replicate the movements and behaviour of people, while simultaneously using a commodity smartphone to create a sensor trace of the movements and behaviours of people using the space. Indicators derived from the replicated sub-population can then be extrapolated to

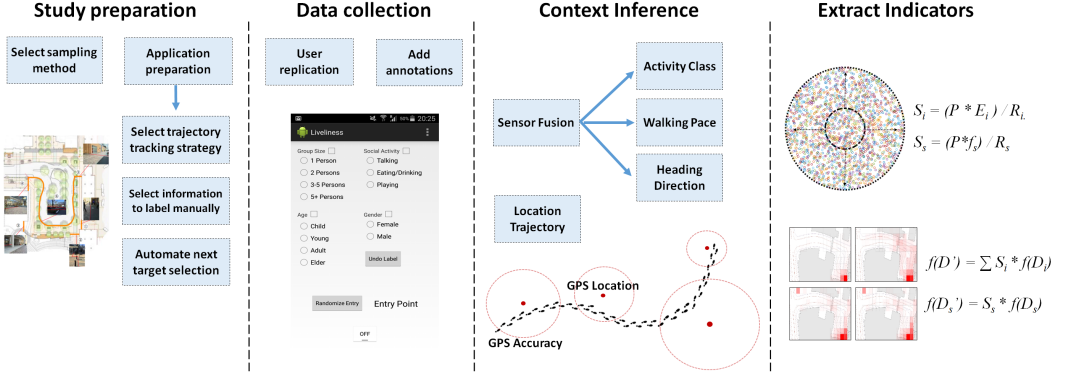


Fig. 1. Overview of the different phases of the crowd replication process.

reflect the behaviour of the overall population. The crowd replication process is summarized in Figure 1 and consists of four main stages. In the following we describe these stages in detail.

Study Preparation: During the first phase of crowd replication, the main tasks are to determine the *sampling strategy*, i.e., deciding the strategy on selecting participants for replication, and to *identify necessary annotation labels* to support visual observation. The sampling strategy should capture a representative subset of the target population, while also ensuring representative sample of population behaviours. Representativeness of the target population can be ensured using a population-based sampling strategy, i.e., replicated users are chosen based on the relevant population characteristics within the target area (e.g., gender, age group or other characteristics). However, this requires prior effort in surveying the population characteristics, which is not always readily available or feasible to conduct. Additionally, this approach does not guarantee the sample to be representative of the behaviour of the people within the space. To simultaneously achieve representativeness of both target users and their behaviours, we rely on *cluster sampling* where the physical entry points of the space are interpreted as clusters and random sampling is applied within each entry point [54]. Our strategy is motivated by the observation that public spaces typically have natural entry points which guide the movements of people in and out of the space, and which thus influence the activities and behaviour of people within the space. Utilizing entry points to select replicated participants causes minimal bias to the distribution of user demographics, while simultaneously allowing to capture complete motion trajectories. In our experiments we demonstrate that this strategy produces data that is representative of the overall target population and their behaviours; see Section 5.1 for details. Note that crowd replication is not restricted to these sampling strategies. Indeed, one of the key strengths of crowd replication is that it enables fine-grained control of the sampling strategy. This control enables studies targeting specific sub-populations, and also mitigates possible biases in the sample population. Such biases are often difficult to avoid and understand with pervasive sensing and opportunistic data collection methods [33].

Besides determining the appropriate sampling strategy, in the study preparation phase the researcher(s) identify labels of relevant activities and demographics that are needed to assess the quality of the target area. These labels are captured through visual observation and complement sensor traces captured through replicated movements. While many physical activities are feasible for direct replication, more complex activities may be easier to capture through direct observation and manual annotation. For example, social activities are easier to capture through visual

observation than through direct replication. However, even for these activities, crowd replication helps to produce more detailed information as the sensor traces captured during replication help to characterize these activities, e.g., how walking characteristics or dwell times are affected by social activities. In contrast to social activities, discerning between optional and necessary activities is more challenging as we do not know the actual motives of the people moving within the area. Certain activities and movement trajectories are more likely to indicate optional activities, e.g., sitting on a bench, or a trajectory between different parts of a shopping center. Similarly, other activities and trajectories are more often associated with necessary activities, e.g., remaining stationary besides a bus stop, or direct path from an entry point to a post office or bank. When choosing annotation labels, requirements for capturing relevant optional and necessary activities should thus be considered. Typically the annotations should include basic demographics about the users, e.g., age and gender of the participant. More detailed annotations can be relevant for certain spaces, e.g., if the participant seems to be shopping, transiting or working. In some cases, such cases can be automatically extracted from the sensor traces. As an example, there are several works on categorizing museum visitors based on their movement patterns [61] and similar techniques could be used to uncover motivations of people using the space. If such information is vital for the study, requesting a brief interview from the participant after the replication process is finished might be a viable approach. In our experiment, only information that can be captured without interrupting the users are considered. Specifically, we consider annotations about participant's age discretised to four categories, participant's gender, three types of common social activity, and group size discretised to four categories.

After the sampling strategy and annotation labels are selected, a mobile application is configured for the area of investigation. The application is responsible for recording sensor measurements and helps with adding manual annotations about direct observations. While both operations can be run on a single device, in practice the best option is to use two devices, one for sensor capture and one for annotation. This helps to ensure the annotation process does not interfere with the sensor traces captured during replication. The device used to collect sensor measurements should be placed and oriented to ensure optimal sensor trace for motion analysis, i.e., in our case we placed the device in upper thigh pocket to ensure we can accurately capture details about each step. As our experiments focus on outdoor spaces, we rely on GPS, magnetometer, accelerometer and gyroscope measurements for recreating the movements and activities of the replicated user. Note that while our experiments do not cover indoor environments, crowd replication could also reasonably be adopted for indoor environments by replacing GPS with a relevant indoor localization technique; see Section 8 for discussion. The interface for adding annotations allows quick insertion of demographics about the participants, as well as labels about activities and behaviours which were not physically replicated by the researcher. This latter option is necessary as replicating some activities can be impractical or cause concern to the replicated person by revealing the replication process and altering their normal behaviour. The exact type and detail of activities to replicate or manually label depend on the objectives of the study, and can be tailored to the mobile app during the study preparation phase. The annotator interface is also used to randomize the next entry point from which to initiate the next replication process. Screenshot of our application is shown in Figure 2a.

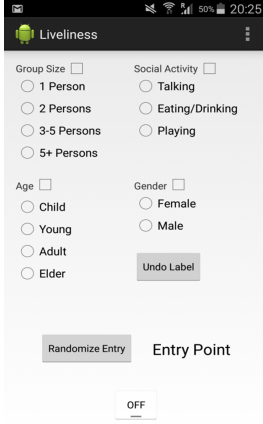
Data Collection: In the second stage, data collection, researchers carrying out the investigation are tasked with observing people moving in the space and replicating their movements and activities on a general level. Each replication is started from one of the entry points randomly selected by the mobile application. The goal is to produce a sensor trace that mirrors the motion of the target user, i.e., the trace captures the motion information the target person would have produced if she had

been wearing the sensing device; see Figure 2b for an illustration. Note that the overall goal of the replication is to capture parameters describing motion on a general level, not to capture the exact sensor representation of walking characteristics. Indeed, we aim at capturing parameters such as approximate position, pace, and step count, which generally can be replicated with relative ease and reasonable accuracy, contrary to exactly mimicking the gait patterns of a user [38]. To collecting a sensor trace with one device, the research uses a second device to insert annotations about visual observations. The application supports quick insertion of user demographics, activities of interest, and other factors of interest. The application automatically links the annotations with the location where they were taken, thus researchers can capture the spatial distribution of activities by making the annotations approximately at the location where they were performed. The overall effort to collect data in crowd replication is moderate and comparable to direct observation, while providing information at finer granularity. For example, in the case study considered in our evaluation (see Section 4), six different researchers collected data for 2 – 3 hours, resulting in a combined data set consisting of roughly 16 hours of measurements.

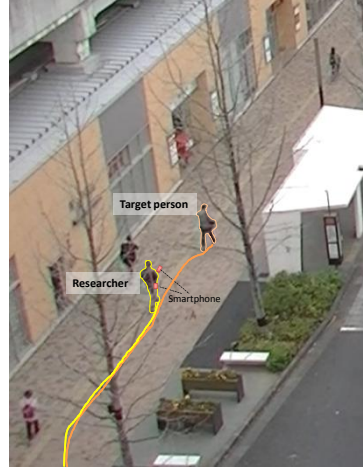
To avoid interfering with normal behaviour within the space, the data collection should be carried out in a subtle manner that avoids making the replication process obvious [23]. In case of a busy area, this typically only requires maintaining sufficient distance to the target person and avoiding such activities which could cause confusion. For example, in our evaluation none of the 614 replicated users showed visible concern or signs of noticing the process. In case of a quiet area, or where following a person is not suitable, the main option is to record user demographics first, and repeat the trajectory only after the user has exited the area. Note that shadowing (or trailing) is widely established as a tool to assess uses of a space. For example, museums often adopt related techniques [6, 13]. Crowd replication can also be combined with other data collection techniques. For example, trajectories can be replicated from a video recording in case such media is available. Note that compared to visual surveillance, requirements for the recording are substantially more relaxed in terms of video resolution, occlusion and lightning conditions. Crowd replication can also be used to supplement video analytics, e.g., by recreating trajectories where visual analytics are unable to capture the measurements sufficiently accurately.

Context Inference: In the context inference stage, the collected sensor data and the recorded annotations are passed to an offline analysis framework. The analysis framework is responsible for reconstructing enriched trajectories containing information about participants location, details about pace, intensity and heading of motion, replicated activities, and recorded annotations. As the first step within the analysis framework, we construct pedestrian models from the motion sensors, i.e., accelerometer and gyroscope, that provide high quality information about step frequency and overall intensity of motion. The pedestrian models are next combined with a heading estimate obtained through a sensor fusion of accelerometer, gyroscope and magnetometer measurements to produce a pedestrian dead reckoning (PDR) system. The PDR system is then fused with available location information from positioning system, i.e., in our experiments GPS, and the coordinates of the initial entry point. The fusion between the PDR and available location information is used to simultaneously manage accumulation of PDR error, and to enhance location accuracy provided by the underlying positioning system, in particular when the location accuracy is low.

Motion sensor measurements are then interpreted by an activity recognition module to identify physically replicated activities. In our experiment we consider identifying between stationary, walking and running. Note that given the freedom to instrument the researcher with any configuration of sensors, more complex set of activities could also be replicated and identified from the sensor measurements; see Section 8 for further Discussion. Finally, annotations about direct observations



(a) User interface of the application used for crowd replication in our example use case.



(b) Example of a researcher carrying out crowd replication.

Fig. 2. User interface of the app we have developed for supporting crowd replication (Fig. 2a) and an example of a researcher replicating the movements of a person (Fig. 2b).

containing details about user demographics and activities not physically replicated are attached to produce final enriched trajectories.

Extract Indicators: The enriched trajectories alone are sufficient for answering many questions regarding the use of public spaces that can be quantified through *relative* differences between usage. As examples, consider evaluation goals such as inspecting fraction of people stopping at an information post or exhibition stand, relative motion flows within the area, or ratio between young and elderly age groups. However, there are also many characteristics that reflect the *total* number of individuals. For assessing these characteristics, it is necessary to generalize the indicators derived from replication to the entire population of the target area. We achieve this by deriving scaling factors that enable adjusting values from replicated data into the overall population. To formalize the extrapolation process, let P denote the total number of people in the target population, E_i the fraction of all users that arrive at cluster i , and R_i the number of users replicated at cluster i . As an example, in our case P corresponds to the total visitor count for a public space, the clusters correspond to different entry points, and E_i corresponds to the apriori frequency of people entering the space through the entry point corresponding to cluster i . The value of P can be estimated from publicly available data sources, such as pedestrian count records, residence counts, or travel survey data, whereas the values of E_i can be estimated from small-scale people counting surveys, e.g., we estimated the values of E_i from two 30 minute video feeds.

Formally, for each cluster i , we derive a scaling coefficient S_i for extrapolation using:

$$S_i = (P \cdot E_i) / R_i \quad (1)$$

Let $f(R_i)$ denote an indicator calculated from the replicated measurements R_i , the adjusted estimator $f(D_P)$ for the entire population P can then be calculated using

$$f(D_P) = \sum_{i=1}^n S_i f(R_i). \quad (2)$$

Note that the scaling factors only take into consideration *active* users for which replicated data is available. In some cases we may be interested in estimators for sub-populations which cannot be reliably replicated (e.g., stationary users, people with baby carriages, and so forth). Estimators for these sub-populations can be calculated using a similar process. Accordingly, we first need an estimate of the apriori frequency of the sub-population, after which manual labels can be used to estimate the number of users in the sub-population, i.e., R_i . The final estimator is then calculated analogously, using Equations 1 and 2.

The length of trajectories could potentially bias crowd replication as long trajectories contain more measurements than shorter ones, and hence are likely to have a larger impact on the scaling coefficients. To mitigate this bias, we normalize the scaling factors with respect to the length of the trajectories, i.e., we derive a normalized estimator $f(D_P^{norm})$ using

$$f(D_P^{norm}) = l \sum_{j=1}^m f(r_j)/l_j. \quad (3)$$

Here $f(r_j)$ is the indicator calculated from replicated trajectory r_j and l_j is the length of the trajectory.

3.2 Pedestrian modelling for Crowd Replication

The quality of the information captured through crowd replication depends on the accuracy of the underlying sensing solution that is used to recreate the activities and movements of the replicated subjects. We have developed a novel high-accuracy sensing solution which, instead of performing online recognition during replication, learns a personalized model of the movement characteristics of an individual from all data collected during replication. By tailoring the sensing to the characteristics of the researcher, higher accuracy can be obtained than using off-the-shelf tracking techniques. In particular, off-the-shelf techniques are vulnerable to changes in pace [7], which are essential for understanding behavioural patterns and capturing social activities. Additionally, as shown by Muaaz and Mayrhofer [38], mimicking another person's walking patterns easily results in asymmetry, which further complicates the pedestrian modelling. In our case we seek to capture the movement trajectory of the subject, and details about his/her walking pace (in steps per second) and pattern in terms of intermediate stops and atypical steps.

Figure 3 shows an overview of our sensing solution. We operate on three dimensional acceleration and angular velocity measurements where acceleration is measured in m/s^2 and angular velocity as radians per second. Following best practices in activity recognition research [5, 22], we divide the measurements into frames of one second. To ensure step cycles are captured as accurately as possible, we use an overlap of $k - 1$ samples where k is the length of the window in samples. To reduce effects of jitter and noise, we pre-process the measurements using a low-pass filter. We then distinguish between periods of activity and inactivity using a threshold on the standard deviation of the overall acceleration [4]. The overall acceleration a is defined as the L^2 -norm of the component-wise accelerations, i.e., $a = \sqrt{a_x^2 + a_y^2 + a_z^2}$. We consider a frame as active whenever the standard deviation exceeds a threshold $\sigma_{active} = 0.30$. This corresponds to a lenient threshold that labels frames with any notable motion as active. We segment the data into periods of activity

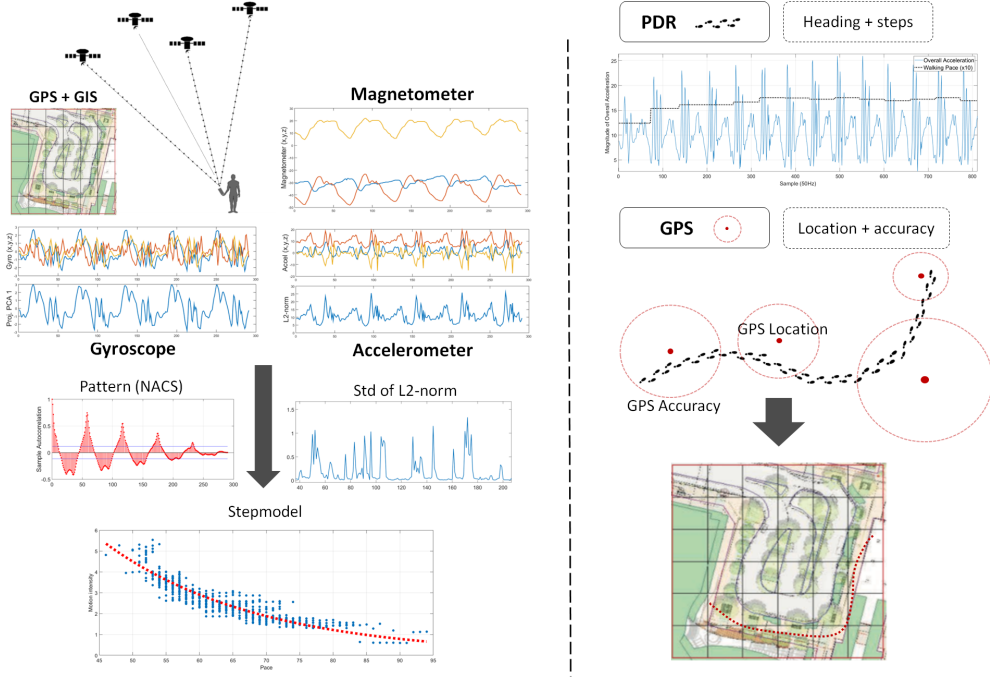


Fig. 3. Overview of mobile sensing approach.

and inactivity by combining successive frames exceeding the activity threshold σ_{active} together. To further improve accuracy, we identify exact start and end times by analysing the first and last frames of each period and considering the first sample to violate the threshold σ_{active} as the boundary point of the corresponding segment.

From the frames labelled as active, we next estimate the walking pace of the user. We consider only gyroscope measurements as we found them to better capture the cyclical patterns associated with walking than accelerometer measurements. Indeed, accelerations caused by heel strike and external activities can mask the dominant pace in accelerometer measurements, while the underlying cycling pattern remains visible in gyroscope measurements. To capture the dominant pace, we find the dominant dimension of the gyroscope measurements by projecting the measurements along their first principal component. From the projected gyroscope measurements, we mine for cyclical patterns to produce (walking) step candidates. To find these patterns, we adaptively divide the data into frames using a grid-search that identifies the optimal frame length and overlap parameters. Specifically, we vary frame length f_t between 1 – 5 seconds, and frame overlap f_o between 0 – 95% to find measurements frames that are optimal in terms of their autocorrelation score. The advantage of our approach is that it can guarantee an optimal framing for the underlying pattern, whereas fixed frame parameters used by most step detection approaches [7] would make the detection susceptible to parameters values that are incompatible with the user's actual walking pattern. Fixed frame parameters would also struggle to identify walking patterns when the underlying walking pace varies.

From each frame we identify autocorrelation peaks ρ_p that exceed a minimum threshold $\rho_{min} = 0.30$ as initial candidate step frequencies. Frames where the minimum threshold is exceeded are then split into sub-frames, referred to as *step frames*, corresponding to sensor measurements for

candidate steps. Autocorrelation lag only provides an estimate of the pace over the entire frame, while duration of each individual steps within the frame can vary. To capture this variation, we find exact step frame boundaries by using minima in overall acceleration as cut-off points. Specifically, we define step frame boundaries by finding local minima from the overall acceleration. To avoid situations where the step duration varies significantly from one step to another, and to account for normal variation in human walking pace, we require these minima to be within 20% of the point corresponding to the initial candidate step frequency. For every detected candidate step, we then store the duration of the step, value of the autocorrelation, minimum, maximum and standard deviation. We note that the proposed approach to instrument the researchers with sensing devices(s) allows controlling many of the most challenging variables in motion sensing systems, e.g., device model, device placement and user characteristics. This is a key benefit allowing robust selection of threshold values when compared to systems that are designed to work in the wild and have to generalize over such variables.

From the mined dataset we learn a regression model between pace and motion intensity, i.e., between duration of candidate steps and standard deviation of overall acceleration during the candidate step. The relation between walking pace and motion intensity is roughly linear for normal walking paces. In case of very slow or fast walk, the relation roughly follows an exponential decay function curve. Typically there are one to three such relationally dependent subsets within the data, each corresponding to a pattern generated by multiple of steps. We learn models for the first two orders as we found these to be the most relevant, while third and consecutive order multiples are redundant to the first two. Indeed, depending on walking pace and device placement, the strongest pattern is produced by either a repetition of a single step or repetition of two steps, often referred to as *stride*. The pattern for a single step is typically stronger when the device is equally affected by steps taken by both legs, e.g., while held in hand. Stride, on the other hand, produces stronger pattern when the device is asymmetrically affected by steps, e.g., when the device is placed within the user's trouser pocket.

To reduce computational complexity of learning the regression models, we cluster frames based on similarity in pace p , motion intensity m and autocorrelation ρ . This is particularly effective during continuous unobstructed walk, which typically exhibits a stable pace and overall motion, producing multiple steps with high similarity. The clusters are formed by processing through the frames in order of descending autocorrelation score. Each sample is tested against existing clusters and added to the model it best fits. In case none of the models provide a good fit (currently we use $d > 0.05$ as threshold) for the sample, we initiate a new cluster with the sample. Formally, distance between sample A and cluster C is calculated as:

$$d = \text{abs}(p_A - p_C)/p_C + \text{abs}(\rho_A - \rho_C)/\rho_C + \text{abs}(m_A - m_C)/m_C \quad (4)$$

After reducing number of candidate steps through clustering, we learn the two regression models that best characterise the candidate steps. This process follows a two-stage approach. First, we initialise two models with seed samples corresponding to the clusters formed around the frame with strongest autocorrelation. We then find feasible candidates that exhibit downward sloping linear relation between walking pace and motion intensity. As there are naturally noise in both step length and the motion intensity values, we allow a degree of slack (currently 5% of the models initial sample) to fulfil the requirement for negative linear relation. We note that at this stage clusters can be valid for more than one regression model.

In the second stage, we produce the final regression models between walking pace and motion intensity. We initialise the two models with all valid clusters for the given model and then follow an iterative process to find clusters which produce most cohesive regression models. During each

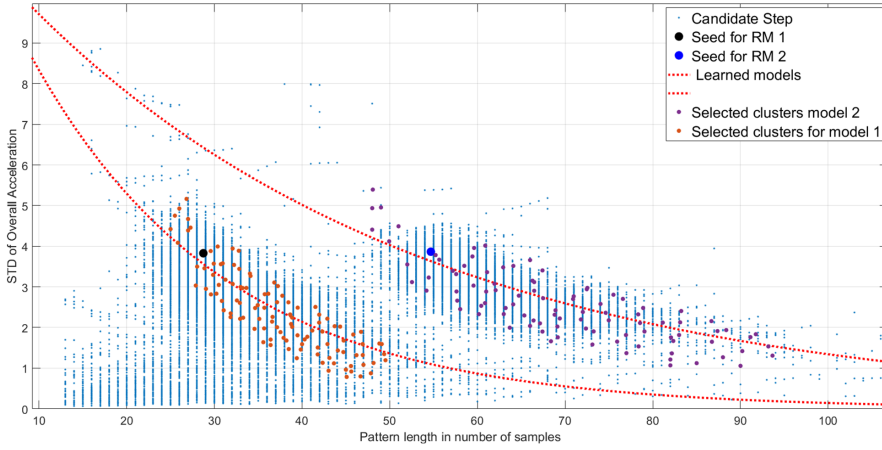


Fig. 4. Regression models discovered from the data between walking pace and std of accelerometer L_2 -norm. In the model $N_1 = 100700$, $RMSE_1 = 0.22$, $R - Squared_1 = 0.8750$, and $N_2 = 203910$, $RMSE_2 = 0.31$, $R - Squared_2 = 0.6536$

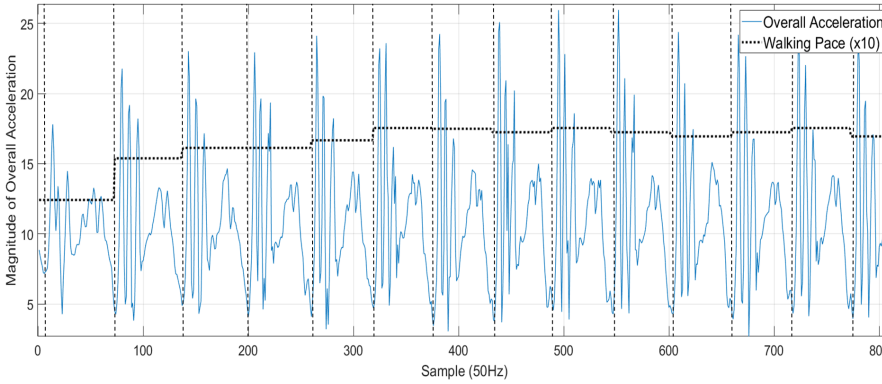


Fig. 5. Illustration of detected walking pace (scaled by 10) aligned with raw data about overall acceleration.

iteration we learn a regression model, and in case the fit is below minimum requirement, remove the cluster with highest residual from the model. To account for the varying number of samples forming the clusters, we use the cluster size as a weight when learning the model. The process is continued until a sufficiently good fit of the data is obtained, or number of remaining data points falls below a threshold. If some clusters remain in both models after the filtering process, we assign each of these conflicting clusters to the model where it produces the smaller residual. The regression models learned from the data are illustrated in Figure 4.

Once the regression models have been learned, candidate steps are validated against the regression model to identify those that exhibit a good fit. We note that there is also an inherited autocorrelation threshold $\rho > .30$ as the candidates are generated only for frames above the threshold. The final outcome in terms of pace aligned over raw motion intensity data is shown in Figure 5.



Fig. 6. Target public space and its nine entry points.

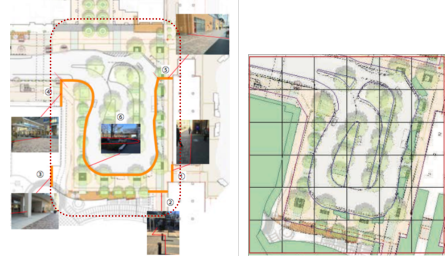


Fig. 7. Center area and its main entry points (left), and the area split into 15x15m grids (right).

3.3 Location Sensing

We recreate movement trajectories of the subjects by combining location information and pedestrian dead reckoning (PDR), augmented with the pedestrian modelling module detailed above. As we initialise each replication from one of the entry points automatically provided and stored by the application, we utilize coordinates of the respective entry point as the initial location. Our experiments focus on an outdoor space, and hence we rely on GPS as our underlying positioning system. For experiments focusing indoor spaces, leveraging an existing indoor positioning technique could reasonably replace GPS; see Section 8. We sample GPS at 1Hz, which ensures high tracking accuracy. Note that, while GPS suffers from heavy battery drain, the sensing application can run continuously for several hours on a contemporary smartphone, which is sufficient for typical data collection in crowd replication. For extended data collection tasks, replacement phones might need to be considered to avoid draining the phone's battery.

To yield final spatiotemporal movement trajectories, we follow a standard fusion of GPS measurements with PDR. Specifically, GPS measurements are used to manage accumulation of error from PDR, whereas PDR is used to improve location accuracy between the GPS measurements, in particular when the GPS reception is low. Our PDR approach is based on detecting steps and pace using the pedestrian modelling module as described above, and a sensor fusion based orientation filtering method proposed by Madgwick [34]. We balance between GPS measurements and PDR estimates by weighting GPS location estimates with horizontal accuracy reported by GPS. Specifically, we allocate the location measurement a weight $\alpha = [0.0, 0.5]$ using a decay function $f(x)$ so that $f(x \geq 15.0m) = 0.0$ and $f(x \leq 5.0m) = 0.5$. Exponential decay is used to determine the appropriate weight for the intermediate values between $x = [5, 15]$ meters.

4 EXPERIMENTAL SETUP

We assess the potential of using crowd replication to gather meaningful data about human behaviour within public spaces through a case study carried out within a representative example of a metropolitan urban space. We carried out three separate crowd replication trials; (T1) a 30-minute trial during weekday with simultaneous video camera recording; (T2) a three hour trial during weekend; and (T3) a three hour trial during weekday. Each trial was carried out by four researchers, each carrying two devices. One device was placed inside trouser pocket and was used to collect sensor traces for replication, whereas the other device was held in the hand of the researcher and used for ground truth annotations about user demographics. The device held in the hand was also used to randomize the next entry point to sample from. T1 took place within the area depicted in Figure 6, while T2 and T3 took place within a sub-area of the overall space shown in Figure 7. In total, the behaviour of 614 individuals was replicated (T1: 86, T2: 228, T3: 300).

4.1 Description of Target Area

We have conducted our field study at the outdoor areas of Kashiwano-ha, which is located within the larger metropolitan area of Tokyo. Kashiwano-ha is at the intersection of a popular shopping centre and a transportation hub encompassing train, bus and taxi stations. We selected this area since it includes various artifacts and spaces that have been designed to improve human experiences. The planners and designers are keen to evaluate the area quantitatively, and recently direct observation and manual pedestrian counting were conducted at several locations within the area. The area is mainly aimed for pedestrians, with cycling prohibited within the area. The area features services such as several sets of tables, total of 95 chairs and benches, information stand, guide map and an interactive touch display. The centre of the area, as portrayed in Figure 7, contains six possible entry points, and the full area, portrayed in Figure 6, contains nine entry points. All data collected through our experiments, as well as the layout of the target environment, have been made available online at http://akiyatoolbox.org/crowd_replication/.

4.2 Crowd Replication

In our experiment, we focus on the outdoor areas at the centre of the Kashiwano-ha. The target area is used for stationary activities including socializing, eating, or spending time, and to move between various locations surrounding the central area, i.e., the two sides of the shopping centre, train and bus stations, hotel, post and bank. We have used crowd replication within our target area to assess the quality of the space and to extract indicators of the quality of the human experience. Below we detail how crowd replication was carried out within the target environment.

Data Collection Design

As our location sampling method, we have opted to use GPS due to clear view of the sky within majority of the target area. The tall buildings near the entry points can cause decrease in localization accuracy, which we compensate by using the coordinates of the entry points to initiate each trajectory. The PDR component of our trajectory tracking approach is valid for the entire area considered, i.e., there are no escalators or moving sidewalks within the considered area. Within the data collection application, we have included manual annotations about:

- Group size of the replicated participant/group,
- gender of the participant,
- age category of the participant, and
- common social activities within the area, i.e., 'Talking', 'Eating/Drinking', and 'Playing'.

Sampling Design

To collect a representative subset of population, we utilized the entry points (six for T1 and nine for T2 and T3) of the area to select the targets for user behaviour replication. Specifically, each new target was selected as the first person arriving from a randomized entry point. Querying for the next entry point was implemented as part of the mobile app we used for sensor data collection. The mobile app was also responsible for recording ground truth about user demographics and social activities. The probability distribution for randomizing the entry points was derived from an earlier survey where we counted the number of people arriving from each entry point. The survey was carried out separately for weekdays and weekends and we used different crowd density distributions for them.

User Behaviour Replication The primary characteristics we focus on replicating include movement and activity related characteristics. In particular, we record the movement trajectory of a subject, his/her physical activities, and any intermittent stationary periods. We also replicate the step length and walking pace of the subject as accurately as possible. In the following section

we demonstrate that these characteristics can be accurately replicated. We focus on periods with movement or physical activities for two main reasons. First, moving users are the most challenging subgroup to survey, while stationary individuals sitting or standing can be efficiently recorded using a combination of pen and paper and mobile tools. Second, we focus our investigation on indicators and user groups that most benefit from the use of mobile sensing.

Model Construction For obtaining aggregate level statistics, we require the scaling factor P for the total number of people visiting the space. We estimated P by aggregating values from three sources; statistics publicly available from the local railway company¹, population count survey recently carried out within our target environment by urban planners and researchers, and our own a video-based population count. Additionally, to obtain sub-population estimates, we performed a small-scale video-based crowd counting to estimate the fraction of stationary and mobile users within the space. The sub-population estimates were derived from a 30 minute video feed that was collected simultaneously with the crowd replication process. Scaling factors for estimators based on active and stationary populations, i.e., $f(D)$ and $f(D_s)$, were then derived using Equations 1-3. Final estimators were obtained by normalizing replicated trajectories by their length and multiplying the estimators with the corresponding scale factors. Note that while visual monitoring was used, we only performed low-effort crowd counting instead of tracking movements of individuals. Crowd counting is low-effort compared to visual tracking and can be carried out also with alternative means, e.g., through direct observation.

5 VALIDITY AND ACCURACY OF CROWD REPLICATION

We begin our evaluation by validating crowd replication as a data collection mechanism. We demonstrate that crowd replication captures a sample that is representative of the key characteristics of people using the target space. Specifically, we demonstrate that both the trajectories collected with crowd replication and the demographics of the sample converge to the underlying distributions of people using the space. Next, we consider the accuracy of the data collected through crowd replication, demonstrating that indicators extracted from replicated measurements closely mirror those captured from manually labelled surveillance data. We also demonstrate the benefits of our pedestrian modelling approach by evaluating it using an existing benchmark dataset [7]. Finally, we critically compare crowd replication to other data collection methods, demonstrating that it can significantly increase sample size compared to direct observation while at the same time requiring significantly smaller resources than visual tracking or other pervasive sensing solutions.

5.1 Sample Validity

The representativeness of the measurements collected using crowd replication depend on how well the overall sample and the replicated trajectories capture key characteristics of the behaviour and demographics of the target population. With the exception of small-scale spaces in homogeneous neighbourhoods, both the behaviours of people and demographics are likely to contain significant variability, making it difficult and non-trivial to accurately capture the underlying distribution of the target population. To illustrate this point, Figure 8 shows the distribution of four key demographics and two key movement characteristics across the different entry points for the larger target area considered in our study (see Figure 6). From the figure we can observe that each entry point differs significantly from every other along at least two characteristics. We can also observe that the entry point distributions differ from the aggregated distribution of the overall population.

We first demonstrate that the trajectories collected with crowd replication converge to the trajectories of the target population. We carry out this evaluation by calculating the minimum

¹Metropolitan Intercity Railway Company <http://www.mir.co.jp/company/number.html> [Retrieved 29-06-2016]

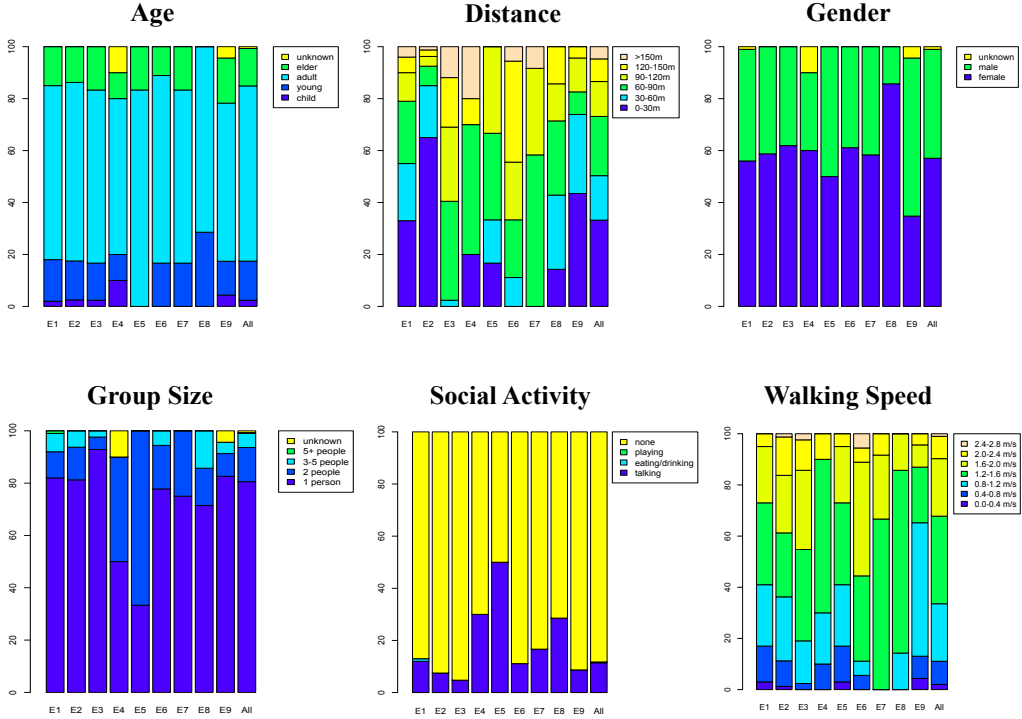


Fig. 8. Distributions of the key demographics and movement characteristics across different entry points. The variables E1-E9 correspond to the nine entry points for the full target area shown in Figure 6.

pair-wise trajectory distance between replicated trajectories and estimate the speed at which the distances converge. Specifically, we iteratively progress through the replicated trajectories and add them to a set of seen trajectories T_S . During each iteration the next unprocessed trajectory is compared to those already in T_S and the minimum distance to a trajectory in T_S is returned. As the distance measure we use the Fréchet distance, which is a widely used measure for assessing the similarity of movement trajectories [19, 51]. In this evaluation, we examine data collected during a three hour data collection session conducted over a regular weekday.

The results in Figure 9(iv) demonstrate that after approximately 50 user replications distance has dropped significantly, indicating that majority of the common trajectories have been mapped. The distance metric continues to decrease through 50 – 300 user replications as increasingly similar trajectories become available for comparison. After 300 replications distance decreases more slowly, indicating a point where majority of movement trajectories and their variations have been observed. We also used a similar methodology to assess the convergence of sample demographics. The results of these evaluations are shown in Figure 9(i)–(iii). From the results we can observe that the probabilities of the different characteristics converge to reflect the actual densities of people using the space after around 50 – 250 replications. In our target space, age category was the most varied taking approximately 250 rounds to converge to underlying population. The fastest category to converge was group size which converged in less than 100 replications. Overall, the results

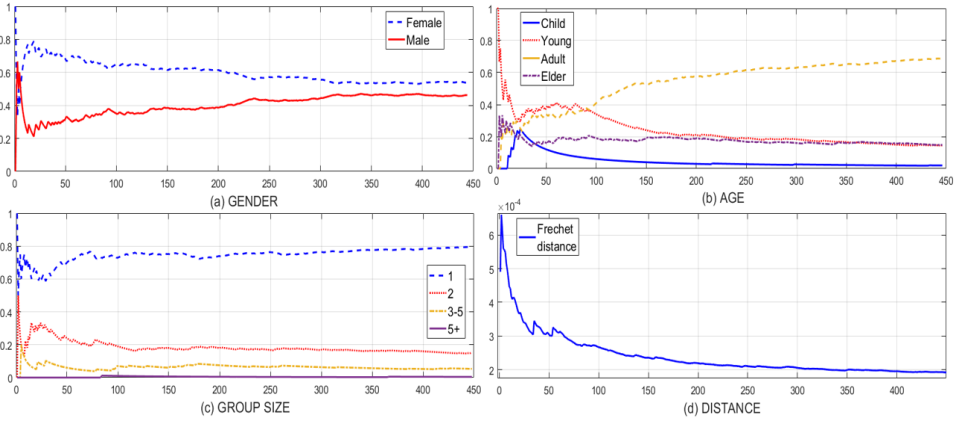


Fig. 9. Number of replicated user trajectories vs. convergence of (i) gender, (ii) age, (iii) group size categories, and (iv) minimum pairwise (Fréchet) distance to the most similar location trajectory.

thus indicate that crowd replication captures a sample that is representative of the behaviours of the people within the space. To put our results into context, the sample size required by crowd replication is comparable to the use of video surveillance which has the highest overall resolution for capturing behaviour. However, the cost of deploying crowd replication is significantly lower than deploying visual surveillance and the data collected through crowd replication can be analysed with significantly lower effort. We also note that the distribution of user demographics were similar for all researchers performing crowd replication, indicating that crowd replication is easy to implement in practice and capable of effectively capturing the actual distribution of users.

5.2 Accuracy of Crowd Replicated Data

As the next step of evaluation we assess the overall accuracy of crowd replicated data. We perform an end-to-end evaluation by comparing indicators extracted from crowd replicated data and manually annotated visual surveillance data. As indicator we consider the *aggregated dwell time*, which corresponds to the aggregate duration of all people passing through an area [36]. According to its definition, people lingering within an area have higher impact on the dwell time than those merely passing through, and hence areas with higher dwell time are livelier than other areas. We carry out the evaluation considering all measurements collected during T1, i.e., a 30 minute weekday trial where both visual surveillance and crowd replication were used to collect data.

We first assess the correspondence of crowd replication and visual surveillance across all users and the entire target space. During the trial crowd replication was carried out by four researchers resulting in approximately 2 hours of data. To compare against visual surveillance, we use the population model described in Section 3 to extrapolate crowd replicated data to cover the entire population. After scaling, crowd replication results in 38 hours 5 minutes and 50.7 seconds of data, while visual surveillance results in 38 hours, 3 minutes and 57 seconds, i.e., a difference of under two minutes or a deviation of 0.08%. Note that scaling is performed separately for each entry point, and hence the result suggest that both our sampling strategy and population model are accurate.

The above result suggests that crowd replication is capable of capturing indicators that represent the behaviour of people within the space as a whole. To better understand how crowd replication performs at finer granularity, we next consider dwell time estimates derived for smaller subsections of the space. For this evaluation, we uniformly partition the target space into 15×15 meter grid

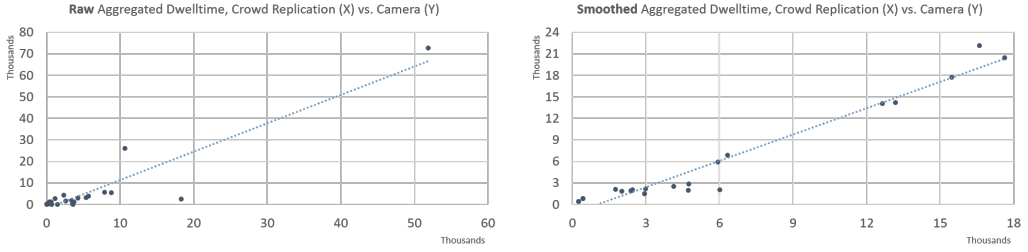


Fig. 10. Comparison of normalized aggregated dwell time between manually annotated video and crowd replicated data, and comparison of spatial distribution of aggregated dwell time.

cells and calculate dwell time separately for each cell. The grid structure we consider is shown in Figure 7. Note that our partitioning is intended as a proof-of-concept for assessing performance of crowd replication, and in practice another kind of partitioning may be more relevant – for example, dividing the space into areas with different functionality. Besides considering dwell time estimates calculated directly from crowd replicated data, we also consider smoothed estimates derived by averaging the values of all cells within one cell’s neighbourhood. Formally, smoothed aggregated dwell time D^s for cell i is calculated using

$$D_i^s = \frac{1}{|N|} \sum_i^N D_i \quad (5)$$

where N contains the (3×3) neighbourhood of i , including cell i , and D_i is the original aggregated dwell time of i . The rationale for considering also smoothed dwell time estimates is that this enables accounting for localization errors in cases where location estimates fall into neighbouring cells, e.g., due to presence of tall buildings.

Figure 10 compares dwell times between crowd replication and visual surveillance. Left-hand side of the figure shows original crowd replicated values, whereas right-hand side shows smoothed estimates. From the figure we can see that both estimates generally align well with visual surveillance. However, we can also observe the original estimates to contain considerably higher errors in some regions. Indeed, highest errors result for the areas located around tall buildings where GPS errors result in dwell times being underestimated. In these cases, the values of the neighbouring cells tend to be overestimated, further exacerbating the dwell time errors. For smoothed estimators we do not observe these errors. Both estimates also correlate well with estimators derived from visual surveillance data, with the original estimates having correlation $\rho = 0.93$ and the smoothed estimates having $\rho = 0.98$. Smoothing also reduces errors across the space as the median absolute deviation of grid cells drops from 1562 seconds to 822 seconds.

To obtain further insights into the errors, Figure 11 shows the difference in dwell time estimates across the space for the two estimators. The estimator derived from original data (left-hand side) suffers significantly around the lower boundary of the area which contains buildings. Also in other areas we can see evidence of location measurements fluctuating across neighbouring cells as regions with underestimated values are next to areas with overestimated values. For the smoothed estimates results are more balanced. The area with the highest error remains the main source of error. However, in other areas the error is almost eliminated. Most errors are within one grid cell, i.e., within 15 meters, which is within the error range of GPS receivers, particularly in the presence of (tall) buildings. Finally, we note that the ground truth we have applied is manually derived and as such can contain minor inaccuracies as perfectly aligning the visual surveillance data with grid

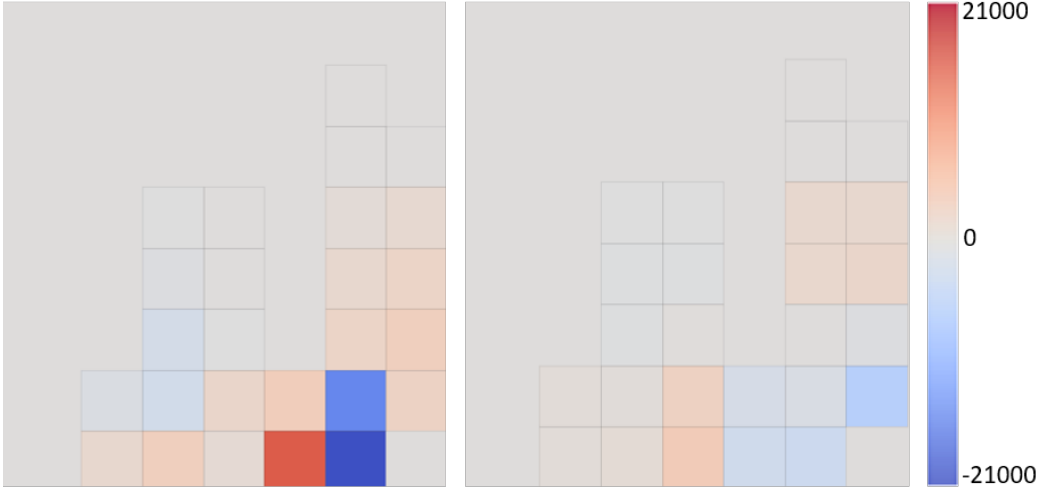


Fig. 11. Comparison of difference between aggregated dwell time derived using visual surveillance and by crowd replication. The left figure contains difference before smoothing, and right figure contains difference after smoothing.

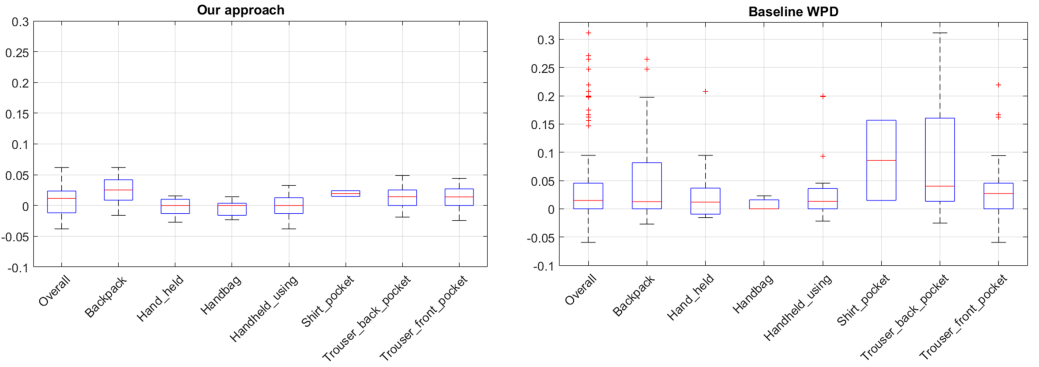


Fig. 12. Boxplots for accuracy of step evaluation in first case, where we only evaluate over known walking periods. The left figure provides results per device placement with our approach, and right figure provides results for baseline with windowed peak detection.

cells is difficult and susceptible to human errors, especially when the user is located close to the border of two cells.

To summarize, crowd replication as a whole is capable of producing accurate and high quality data about human behaviour. However, when finer spatial granularity is required, crowd replication can suffer from some localization error. In our experiments, most errors were within one grid cell, i.e., 15 meters and thus reasonable for practical purposes.

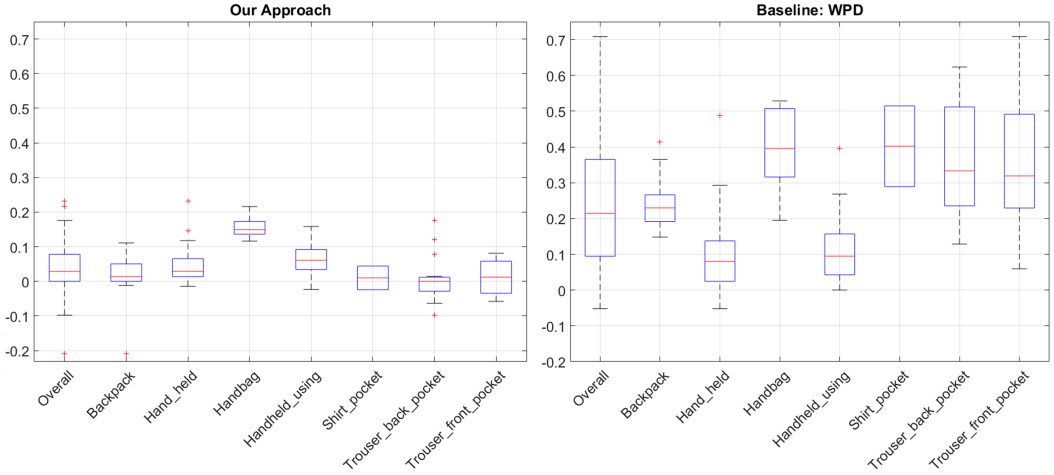


Fig. 13. Box plots for accuracy of step evaluation in second case, where we evaluate entire trace and rely on walking detection to identify walking periods. The left figure provides results per device placement with our approach, and right figure provides results for baseline with windowed peak detection.

5.3 Accuracy of Pedestrian Modelling

We next assess the accuracy of our solution for pedestrian modelling and demonstrate that it outperforms current state-of-the-art techniques. As the trajectories collected from our experiments do not contain ground truth step counts, we carry out the evaluation using a benchmark dataset collected by Brajdic and Harle [7]. The dataset contains accelerometer measurements from walking behaviour which have been collected from 117 trials with 21 users. Ground truth is provided for 102 of these traces. The dataset covers 7 typical device placements and each trial features a period of fast, normal and slow walk. As the data only contains accelerometer measurements, we modify our pedestrian modelling by substituting gyroscope with accelerometer measurements smoothed with a centered moving average filter with a 10ms window length. For other parts our pedestrian modelling remains unchanged and follows the steps described in Sec. 3.2. We separately consider step recognition accuracy with and without walk detection. In the former case we consider traces that have been labelled as walking, and in the latter case we consider the full benchmark dataset. To provide a baseline for evaluation, we implemented the best performing techniques from the paper of Brajdic and Harle [7]. For step counting we implemented a windowed step detection algorithm, and for walk detection we used a threshold on overall acceleration (L2-norm). The best performance for the windowed step detection was obtained with parameter values $\text{MovAvr}_{\text{win}} = 0.28\text{s}$ and $\text{PeakWin}_{\text{win}} = 0.4\text{s}$.

Results for the walk segments are shown in Figure 12. Median error rate of our technique is 1.7% across all walking periods, with the highest errors resulting when phone is placed on the upper body or a backpack. When the device is held in hand or placed in trouser pocket, which would be the most typical placements for collecting data in crowd replication, the errors are consistently below 1%. In contrast, median error of the baseline is 2.3%. The baseline performs slightly better for handbag and backpack placements where the signals are generally symmetric, but has high error rate for trouser pocket where the signal is likely to be asymmetric. As shown by Muaaz and Mayrhofer [38], mimicked walking behaviour often contain asymmetry, implying that our

	Work Effort			Sample	Data resolution	Bias	Privacy risks	Data richness	Recommended for
	Preparation	Data	Analysis						
Direct Observation	small	moderate	large	small	fine grained	small	moderate	rich	Qualitative research (w/ demographic & activity data)
Visual surveillance (manual coding)	large	moderate	very large	large portion in small area	fine grained	small	high	poor	Quantifying short trips in small spaces (exploratory analysis)
Visual surveillance (computer vision)	large	moderate	very small	large portion in small area	fine grained	small	high	poor	Quantifying short trips in small spaces
Voluntary location tracking (GPS, GSM or CDR)	large	very small	very small	small portion in large area	less coarse	mobile users	small	poor	Quantifying city-wide journeys
Location surveillance (GSM or CDR)	small	very small	very small	good portion in large area	coarse	mobile users	high	poor	Quantifying city-wide journeys
Crowd sensing (wifi, bluetooth, etc.)	moderate	small	very small	large portion in medium area	less coarse	mobile users	high	poor	Quantifying trips in large indoor spaces
Crowd replication	small	moderate	very small	large portion in small or medium area	fine grained	small	moderate	rich	Quantifying trips in neighborhood-scale spaces encompassing multiple streets (w/ demographic & activity data)

Fig. 14. Comparison of data collection methods.

technique is likely to perform even better for replicated data than the baseline. Current state-of-the-art step recognition approaches remain highly effective for *online* detection. However, in crowd replication all measurements can be analysed offline, which enables our approach to improve on step recognition performance.

The results across all segments, i.e., when step recognition needs to be combined with walking detection, are shown in Fig. 13. The error rate of our technique remains low, with handbag placement being the sole exception. Baseline performance, however, is significantly impacted by a high number of false positives. In particular, the baseline walk detection achieves only 21.5% accuracy, which would affect the total step count. Our approach is only slightly impacted by additional data, obtaining median overall accuracy of 4.1%, demonstrating the benefits and need for our pedestrian modelling approach. Indeed, the error rate of the state-of-the-art algorithms would significantly overestimate the amount of walking, and thus result in inaccurate indicators and activity estimates.

5.4 Comparison of Data Collection Effort

Crowd replication has been designed as a cost-efficient data collection mechanism that combines direct observation, mobile sensing and mathematical modelling in a way that aims at overcoming the primary limitations of the individual techniques while preserving their main primary advantages. To demonstrate that crowd replication indeed achieves this, we next critically compare different data collection mechanisms.

Figure 14 compares different approaches for data collection in terms of their work rate, effective sample size used for analysis, resolution of collected data, potential biases in the data collection, richness of the collected measurements, and severity of privacy risks. For work rate we further consider the level of effort in different phases of the study. The work effort for study preparation is generally larger for approaches requiring installation of appropriate equipment, recruiting volunteers, and obtaining necessary permissions. Compared to direct observation, our approach requires some additional steps in study preparation phase, as detailed in Sec 3. While the work effort is slightly larger compared to direct observation, it remains small compared to visual surveillance, voluntary location tracking or setting up a crowd sensing approach. The data collection phase in crowd replication requires moderate effort as data is collected by replicating each participant. Compared to visual observation, the work effort is still smaller since researcher's location is automatically tracked and the mobile application interface facilitates inserting labels. Once correctly set, voluntary location tracking and location surveillance require small effort as data is mostly collected automatically. The data collection effort of visual surveillance is comparable to direct observation, requiring monitoring of the equipment and labelling any obstructions, e.g., camera becoming occluded due to traffic or severe weather. Work effort for analysis of the collected data is generally small across all considered approaches, with the exception of direct observations and manually decoded visual surveillance. In particular manually decoding the recorded video(s) can entail an almost overwhelming work effort. For example, detailed annotation of the 30 minute feed from our experiment required over 50 hours of manual labour and only produced aggregate level information of people using the space. Crowd sensing entails similar overall work effort with crowd replication, but suffers from potentially severe biases as only the subset of users with compatible technology (e.g., have a smartphone with WiFi or Bluetooth turned on) is captured, and no controlled sampling design can be applied. Additionally, crowd sensing entails high privacy risks and granularity of collected data is often coarse, making it difficult to obtain a detailed understanding of the urban space. Finally, location surveillance through cellular records has overall the smallest work effort, but can only capture information at coarse granularity.

The sample size for crowd replication depends on the number of researchers employed for the replication task, and the overall number of people using the target space. Generally the replication process yields a small-to-medium sized core sample, which can then be generalized to total population using the scaling coefficients detailed in Sec 3. Automated visual surveillance seemingly provides the best performance, being capable of capturing large sample size with fine-grained information. The main limitations of automated visual surveillance are its privacy intrusive nature, challenges in setting up visual surveillance, and cost and expertise required to operate it. Specifically, cameras have to be installed at locations that provide a clear view of the entire target space. Static objects such as trees, walls, and other physical structures can block the view from cameras, often requiring compromises between quality of recording and practical constraints about viable installation locations. Additionally, partial occlusion due to dynamic objects such as large vehicles (e.g., buses) is hard to avoid or predict. For voluntary tracking the resulting sample is necessarily biased to users willing to participate in the data collection. As such, the method is best suited for studying specific sub-populations with similar behavioural characteristics. For example, voluntary tracking is well-suited for analysing behavioural patterns of tourists. A traditional visual observation relying on pen-and-paper can arguably record activities more precisely compared to crowd replication. The researcher equipped with pen and paper is less constrained compared to the fixed labels provided by the annotations application. The use of fixed labels however is a matter of research design, and free-form text, drawings or similar input methods can also be supported on mobile devices. Finally, direct observation can produce sample bias as participants are selected by



Fig. 15. Spatial distribution of aggregate dwell time during weekdays for non-social activities (left), talking (middle), and dining (left).

the researcher, whereas crowd replication relies on the cluster sampling method to minimise the selection bias.

A central challenge with automated surveillance is that, while performance is generally good, it is not perfect. Indeed, automated surveillance can lead to errors which are particularly hard to observe or avoid due to the automated nature of the process. Manual annotation could be used to support the automated analysis, but this can lead to overwhelming work effort. In contrast, collecting the samples for crowd replication took 2 – 3 hours for four researchers (combined effort of 16 hours). As a hybrid option, existing techniques could well be used to complement crowd replication by automatically estimating total population P or for obtaining scaling coefficients. As an example, in our test-case we combined direct observation about inactive (stationary) population with crowd replication data about active population, effectively capturing both subgroups with most time and resource efficient manner. Crowd replication can also be used to complement other techniques by providing finer detailed samples of the overall population flows, which could then be used as core data points for interpreting rest of the data.

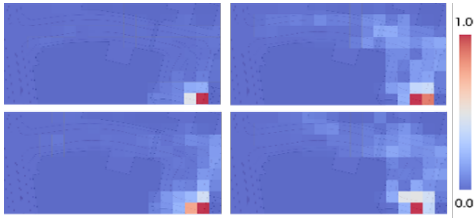
To summarize, crowd replication is able to combine benefits from more resource and labour intensive methods, while requiring only commodity devices and a few active volunteers. Moreover, crowd replication lends itself for combined designs, thus providing a cost effective and easy-to-implement technique that can be used as standalone evaluation method or combined with other methods when even larger sample size or finer data granularity is required.

6 QUANTIFYING HUMAN BEHAVIOURS

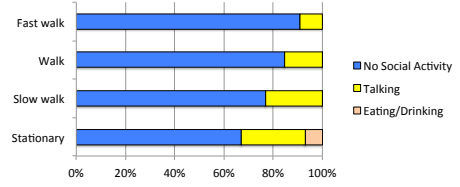
The results of our evaluation thus far have validated the use of crowd replication as a data collection mechanism and demonstrated that it captures data that is representative of the key demographic characteristics and the considered activities of the target population. Next, we demonstrate how crowd replication can be used to extract indicators for assessing public space quality. Specifically, following Gehl's categorization of activities within public spaces into necessary, optional, and social [17], we demonstrate that crowd replication captures spatiotemporally fine-grained indicators of all of these categories. To achieve this, we consider data from our three-hour trials T2 and T3 that took place within the central part of the overall area; see Section 4 for description of the area and Figure 6 for an illustration.

6.1 Fine-Grained Quantification of Activities

We first demonstrate that crowd replication captures spatiotemporally fine-grained information about activities taking place within a public space. By relating these details with specific activity categories, it is possible to characterize and understand how well the space supports the activities.



(a) Comparison of the spatial distributions of aggregated dwell time during weekdays (left) and weekends (right), and of slow walk (upper) and normal walk (lower).



(b) Ratio of different motion intensities i.e., fast walk, walk, slow walk, and stationary.

Fig. 16. Quantified fine-grained activities.

Indeed, presence of specific activities can suggest particular usage patterns and hence is essential for quantifying the human experience of public spaces. Also, diversity of activities is often considered important for a public space to be successful.

In our empirical evaluation, researchers used the mobile application to annotate observed activities such as talking, eating and playing, and subsequently recorded trajectories of the selected activities by following the crowd replication process. The data we collected thus allows us to map the spatial distribution and characteristics of social and other activities.

Figure 15 compares the spatial distributions of different activities during weekdays. Through correlation analysis, we found that the ratio of social activity in grid cells is correlated with group size very strongly during both weekdays and weekends, with $r = .830$, $p < .001$ and $r = .809$, $p < .001$, respectively. On weekdays, they are also moderately correlated with stride duration with $r = .450$, $p < .001$ and weakly correlated with the variance of heading with $r = -0.263$, $p < .005$. Social activities can occur more frequently in a space where people walk more slowly with larger stride duration, and their direction of movements are more predictable. These results reinforce the common understanding that people walk slowly and linger about in socially vibrant spaces. However, thus far such observations have been based on qualitative measurements. The integration of sensing as part of crowd replication allows us to gather also quantitative evidence to support these kinds of observations.

We next examine the potential of predicting the presence of social activities for each trajectory. We first used logistic regression to assess the effect of group size and duration of a trajectory on the likelihood that the trajectory involves social activities during weekdays. The model explained 50.3% (McFadden R^2) of the variance in social activities and a 10-fold cross validation showed that the model correctly predicted 91.3% of cases. Secondly, we performed similar logistic regression to ascertain the effect of group size and motion intensity on the likelihood of social activities during weekends. The resulting model explained 81.8% (McFadden R^2) of the variance in social activities and a 10-fold cross validation showed model correctly predicted 97.8% of cases. The larger the group size and the longer the time length of a trajectory, the more likely social activities such as talking, eating/drinking, and playing take place. However, larger motion intensities were associated with reduced likelihood for the occurrence of social activities; see Figure 16b. Thus, the details captured by crowd replication not only help to understand space usage better, but can be used to predict different activities within the space.

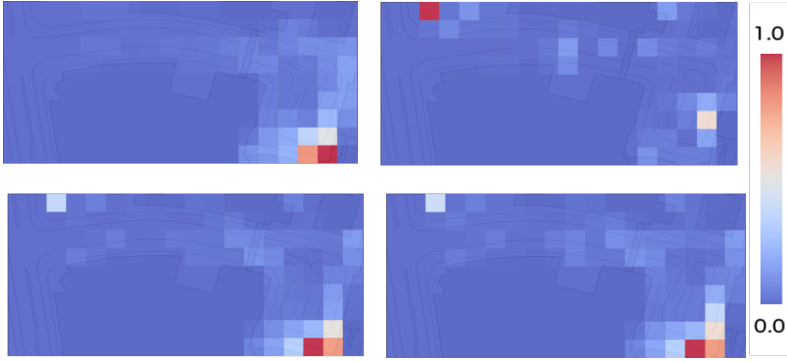


Fig. 17. Spatial distribution of the amount of mobile (upper left), stationary (upper right) and social activities (lower left), and a liveliness index (lower right) during weekdays.

6.2 Liveliness of Public Spaces

Liveliness, defined as a product of the number of people and the duration of their stay [17], is often used as an overall measure of the activity taking place within a space. We next demonstrate how crowd replication can be used to characterize overall liveliness, and to provide fine-grained details about spatial and temporal variations in it. We derive an *index of liveliness* by computing the total amount of stationary and mobile social activity within a space [36]. Figure 17 shows spatial distribution of the amount of mobile, stationary and social activities and the resulting liveliness index during weekdays. The figures show the total amount for the 3-hour trial. Alternatively, we can aggregate data for shorter time windows to examine the spatiotemporal dynamics of the public space. As expected, the target area is livelier during weekends than weekdays with liveliness indices of 308,132 and 554,865, respectively.

6.3 Quantifying Fine-Grained Motions

The use of mobile sensing in crowd replication enables collecting detailed motion information as a part of the recorded trajectories. This makes it possible to analyse patterns that are difficult to quantify using direct observation. As a simple example, we compared the mean motion intensity and walking pace between weekdays and weekends. Both the intensity (mean 2.75 and 2.57) and walking pace (mean 1.03 and 1.01 steps per second) were lower during weekends than weekdays. While such findings are intuitive and perhaps not surprising, making this kind of observations is labour-intensive and non-trivial using direct observation or visual surveillance, whereas they can be easily identified with the help of crowd replication. Note that capturing these differences does not require exactly mimicking stride patterns of the target users, but results naturally from trajectories of similar length having different duration.

Finally, we compute the spatial distributions of normalized aggregated dwell time for different motion intensities. Figure 16a shows the spatial distributions of aggregated dwell time during weekdays and weekends. The correlations show that normal walk during weekdays and slow walk during weekdays are the most similar with $r = .94, p < .01$, and that slow walk during weekdays and normal walk during weekends are least similar with $r = .63, p < .01$; As another example, Figure 18 shows the spatial distribution of stride duration for weekdays and weekends. From the figure we can see more variation in the stride patterns during weekdays with the shortest strides taking place around building exits. During weekends, stride duration are more uniform across the space, but overall slightly longer than during weekdays.

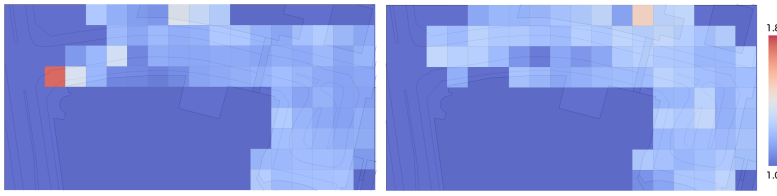


Fig. 18. Comparison of the stride duration during weekdays (left) and weekends (right).

7 DISCUSSION AND FUTURE WORK

Challenges with Replication Process Crowd replication is best suited for capturing people moving in the space as prolonged shadowing could result in the replicated people becoming aware of the process. Stationary people within the area might also become aware of the process by noticing the same people going repeatedly back-and-forth in the area. Potential ways to mitigate these issues are to limit the number of users that are replicated at a given time and to carry out the data collection over a longer period of time (e.g., 30 minutes across several days), or to periodically swap the researchers responsible for replicating users. In some cases, tracking a participant might need to be aborted to avoid revealing the replication process. Generally, when only few people are using the space and/or they are mostly stationary, it is better to use an alternative technique, such as paper annotations or video recording combined with post-hoc replication, to capture the data.

Quality of Crowd Replicated Data The results of our evaluation (see Sec. 5 and Sec. 6) demonstrated that crowd replication captures representative and fine-grained information about the movements and activities of people within a public space. The main source of errors in crowd replication is inaccurate positioning, which can result in errors in the spatial mapping of movements and activities. For example, in our evaluation, areas close to tall buildings resulted in increased location error (see Section 5.2). Indeed, quality of the spatial mapping is highly dependent on the accuracy of the underlying positioning system. When localization accuracy suffers, crowd replication can still be used to assess behaviour of people using the space as a whole, but fine-grained spatial information may be impossible to obtain.

Positioning techniques While our trials have focused outdoor environment where GPS provides readily available location information, many interesting application areas reside indoors, prohibiting the use of GPS. In such environments an indoor positioning system can provide means for inserting location information to the crowd replication system. State-of-art indoor positioning systems commonly rely on particle filters and back-propagation to resolve complete trajectories, making them particularly well suited for offline analysis of sensor traces, i.e., they can effectively leverage information from complete traces that crowd replication provides. Furthermore, the design of crowd replication can help alleviate some of the main challenges for indoor localization by providing exact start and end locations, and by allowing optimized sensor placement to accurately perform step detection and heading estimation. State-of-the-art indoor pedestrian dead reckoning solutions are capable of achieving location accuracy comparable to outdoor environments [15, 42]. Hence, it is reasonable to assume that when leveraging additional benefits from the crowd replication design, the localization accuracy is able to provide fine-grained spatial information also in indoor environments. However, we recognise that such cases need to be validated through future experiments.

Other Application Areas Crowd replication has been primarily designed as a data collection method for urban designers, architects and other stakeholders that are interested in assessing

whether a given space meets its design goals. However, crowd replication is not limited to public spaces, and can be adopted to other spaces. For example, the crowd replication process could be used in museums, malls, supermarkets, airports, or train stations – either as standalone data collection technique or in combination with other techniques.

Supporting Qualitative Studies The focus of this paper has primarily been on capturing quantitative data about public space usage. However, crowd replication can also support qualitative studies. While carrying out the replication process, the researcher experiences the public space and becomes aware of subtle things affecting users behaviours and obtaining a more thorough understanding of the user experience provided by the space. As another example, ethnography researchers could use crowd replication to create a digital trace of their experience, which can then later be used as a memory cue and to reflect on the researcher's experience of the space.

Privacy and Ethical Concerns Recording only sensor measurements instead of other mediums, e.g., video or audio information [40], has the added benefit of increasing privacy of the tracked users. The most immediate advantage is that this information cannot be used to identify users. Another privacy-related advantage of crowd replication is that we are ultimately recording measurements about the researcher instead of the target user. However, we acknowledge that the process of user behaviour replication should be carried out in a way that avoids any privacy or ethical concerns, as well as avoids any socially awkward situations to either target user or the researcher performing the replication. The necessary constraints depend on the characteristics of the target space, but should be planned so that the user behaviour replication can be conducted without the target user becoming aware or disturbed by the replication. This is also crucial for reducing possible observer biases in the collected sample. As an example, we kept approximately 10-15 meters distance between the researcher and the target user and aborted the process in case the target user seemed to become aware of the researcher. In areas where direct replication is unsuitable, replication can be carried out with the help of a video recording from the area. In such case targets users are selected and replicated based on observing their trajectories and activities in the recording. We note that, while this is not the ideal in terms of sampling, it still carries two advantages compared to automated or manual video decoding. First, the requirements for video are much lower compared to those in automated video surveillance. Second, the movement trajectory contains richer information in both granularity and detail than what manual annotation can provide. Also, some details may be difficult for video surveillance, such as determining stride length or intensity. Finally, physically replicating the target users trajectory and activities in the actual environment allows better insight into the human experience of the space.

Scalability Our experiments were carried out in a moderately-sized metropolitan urban public space (the central area is approximately 40×80 meters and contains six entry points). Replicating users within this space was reasonable in terms of work effort, and overall feasible to carry out. In particular, the flow of people within the space was sufficiently high to ensure target users were not aware of the replication process, and the motion trajectories were of reasonable length to limit physical and cognitive efforts of the researchers. For very large spaces, e.g., ones spanning several hundred meters, crowd replication may be unsuitable. Firstly, shadowing another person for a prolonged period of time is prone to making the user aware of the replication process. Additionally, the researcher carrying out the replication may start to lose concentration, which could result in inaccurate data (e.g., decrease quality of mimicking or result in erroneous or missing labels). Secondly, while errors in our location and pedestrian tracking are very small, they accumulate over time and hence data quality would decrease. A potential solution for such spaces would be to divide the area into sub-regions and to perform crowd replication within these sub-regions.

Alternatively, larger spaces could benefit from a combination of different techniques, including crowd replication. Note that the also other data collection techniques face difficulties in very large spaces. For example, visual surveillance would require sufficient unobstructed camera coverage of the space, which in most cases is highly difficult to arrange.

Future Work As one of the main future directions, we plan to investigate combining crowd replication with other techniques for evaluating public spaces. In particular, we look forward to testing a hypothesis that crowd replication can provide validation and correction data to enhance quality of passively collected (e.g., using crowd sensing) measurements about user behaviour within target area. We also plan to investigate performance of crowd replication in a commercial indoor environment with existing indoor positioning system available, and performing crowd replication from a prerecorded video. In the former scenario our main interest is in validating the ability to provide fine-grained spatiotemporal data within indoor environments, while in the latter case we plan to replicate more complex user activities and investigate automatically identifying these from the sensor measurements.

8 SUMMARY AND CONCLUSION

We have contributed by developing crowd replication as a novel and cost-effective mechanisms for collecting spatiotemporally fine-grained data about human behaviours within public spaces. Crowd replication builds on a novel high accuracy motion tracking approach that learns a personalized model of the movement characteristics of the researcher performing replication. Experiments carried out within a metropolitan urban space demonstrated that the data collected with crowd replication closely mirrors the actual behaviour of crowds within public spaces and enables spatiotemporally fine-grained quantification of the considered activities within the target space. As one of the original design goals, crowd replication is easily adoptable and requires no installation of specific hardware, recruitment of outside volunteers, or acquiring specific permissions. Indeed, high-quality data about the use of target space can be acquired using off-the-shelf contemporary smartphones and few hours of ground work from a small group of active researchers.

Crowd replication is advantageous as a data collection mechanism as it enables generalizing to a large population without suffering from privacy and sample bias – two main issues hampering current opportunistic pervasive and mobile sensing techniques. As crowd replication allows collecting samples from all people in a public space regardless of whether or not they own mobile phones (or have specific technology, such as WiFi or Bluetooth switched on), its sampling bias is minimal. Compared to traditional direct observation, crowd replication provides two key benefits: (i) mitigation of selection bias by leveraging a sampling strategy to select participants for the study, and (ii) enhancing direct observations with passively collected sensor traces, providing automatically collected spatiotemporal and motion related context to the observations. Compared to opportunistic sensing which focuses on the scales of cities and greater regions, crowd replication can collect more finer-grained data in terms of spatial resolution, and more detailed data in terms of recorded activities and motion information. Compared to visual surveillance, crowd replication can cover a relatively large and complex area that encompasses streets, neighbourhoods, and occluded areas which would be impractical to sufficiently cover with cameras. Moreover, crowd replication can produce richer data as it can capture subtle patterns of human motions and activities together with location information. Finally, crowd replication can also be easily integrated with background information, such as demographic and subjective questionnaires, making it a versatile and attractive data collection method for understanding and evaluating public spaces.

Our trials took place in an outdoor public space located near a local transportation hub. However, crowd replication is a generic method that can be adopted for other public spaces, including malls,

airports and parks. We encourage that crowd replication can also be adopted for different purposes, including comparisons across different demographic attributes and evaluations, e.g., focusing on baby carriages or tourists, and for different cultural contexts and social norms. Doing so can involve uses of different sampling rates and strategies, as well as different positioning techniques and activity monitoring solutions. This would enable evaluation of the impact of crowd replication on different urban planning practices. Researchers may incrementally form and update the list of activities to annotate and capture in their crowd replication environments by employing iterative research processes. The resulting data from our experiments has been made freely available for research and academic purposes at http://akiyatoolbox.org/crowd_replication/.

REFERENCES

- [1] Alexandre Alahi, Kratarth Goel, Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, and Silvio Savarese. 2016. Social LSTM: Human trajectory prediction in crowded spaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 961–971.
- [2] Miho Asano, Takamasa Iryo, and Masao Kuwahara. 2010. Microscopic pedestrian simulation model combined with a tactical model for route choice behaviour. *Transportation Research Part C: Emerging Technologies* 18, 6 (2010), 842–855.
- [3] Michael Batty. 2001. Agent-based pedestrian modeling. (2001).
- [4] Sourav Bhattacharya, Henrik Blunck, Mikkel Kjærgaard, and Petteri Nurmi. 2015. Robust and Energy-Efficient Trajectory Tracking for Mobile Devices. *IEEE Transactions on Mobile Computing* 14 (2015), 2.
- [5] Sourav Bhattacharya, Petteri Nurmi, Nils Hammerla, and Thomas Plötz. 2014. Using unlabeled data in a sparse-coding framework for human activity recognition. *Pervasive and Mobile Computing* 15 (2014), 242–262.
- [6] Sophie Boonen, Martijn van der Heijden, and Elisa Giaccardi. 2018. Design Is Not for Us: Engaging a New Audience for the Design Museum by Changing Their Expectations. In *Museum Experience Design*. Springer, 45–74.
- [7] Agata Brajdic and Robert Harle. 2013. Walk Detection and Step Counting on Unconstrained Smartphones. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 225–234.
- [8] Francesco Calabrese, Laura Ferrari, and Vincent D. Blondel. 2015. Urban sensing using mobile phone network data: a survey of research. *ACM Computing Surveys (CSUR)* 47, 2 (2015), 25.
- [9] Joshua Candamo, Matthew Shreve, Dmitry B Goldgof, Deborah B Sapper, and Rangachar Kasturi. 2010. Understanding transit scenes: A survey on human behavior-recognition algorithms. *IEEE transactions on intelligent transportation systems* 11, 1 (2010), 206–224.
- [10] Yu Cao, Daniel Barrett, Andrei Barbu, Siddharth Narayanaswamy, Haonan Yu, Aaron Michaux, Yuewei Lin, Sven Dickinson, Jeffrey Mark Siskind, and Song Wang. 2013. Recognize human activities from partially observed videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2658–2665.
- [11] Stephen Carr, Mark Francis, Leanne Rivlin, and Andrew Stone. 1992. *Public Space, Environment and Behavior Series*. Cambridge University Press.
- [12] Nicola Dempsey. 2008. Quality of the built environment in urban neighbourhoods. *Planning, Practice & Research* 23 (2008), 249 – 264.
- [13] Judy Diamond, Michael Horn, and David H Uttal. 2016. *Practical evaluation guide: Tools for museums and other informal educational settings*. Rowman & Littlefield.
- [14] Piotr Dollar, Christian Wojek, Bernt Schiele, and Pietro Perona. 2012. Pedestrian detection: An evaluation of the state of the art. *IEEE transactions on pattern analysis and machine intelligence* 34, 4 (2012), 743–761.
- [15] Moustafa Elhamshary, Moustafa Youssef, Akira Uchiyama, Hirozumi Yamaguchi, and Teruo Higashino. 2016. Transit-Label: A crowd-sensing system for automatic labeling of transit stations semantics. In *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 193–206.
- [16] Jan Gehl. 1977. *The interface between public and private territories in residential areas: a study by students of Architecture at Melbourne University*. Technical Report. Department of Architecture and Building. Melbourne: Melbourne University.
- [17] Jan Gehl. 2011. *Life between buildings: using public space*. Island Press.
- [18] Ensiyeh Ghavampour, Mark Del Aguila, and Branda Vale. 2017. GIS mapping and analysis of behaviour in small urban public spaces. *Area* (2017).
- [19] Joachim Gudmundsson, Patrick Laube, and Thomas Wolle. 2011. Computational movement analysis. In *Springer handbook of geographic information*. Springer, 423–438.
- [20] Mordechai Haklay, David O'Sullivan, Mark Thurstain-Goodwin, and Thorsten Schelhorn. 2001. So go downtown: Simulating pedestrian movement in town centres. *Environment and planning B: Planning and design* 28, 3 (2001), 343–359.

- [21] Samuli Hemminki, Keisuke Kuribayashi, Shin'ichi Konomi, Petteri Nurmi, and Sasu Tarkoma. 2016. Quantitative Evaluation of Public Spaces using Crowd Replication. In *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*.
- [22] Samuli Hemminki, Petteri Nurmi, and Sasu Tarkoma. 2013. Accelerometer-Based Transportation Mode Detection on Smartphones. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems (SenSys)*. ACM, 13:1–13:14.
- [23] M.R. Hill. 1984. Stalking the Urban Pedestrian: A Comparison of Questionnaire and Tracking Methodologies for Behavioral Mapping in Large-Scale Environments. *Environment and Behavior* 16, 5 (1984), 539.
- [24] Serge P Hoogendoorn and Piet HL Bovy. 2004. Pedestrian route-choice and activity scheduling theory and models. *Transportation Research Part B: Methodological* 38, 2 (2004), 169–190.
- [25] Weiming Hu, Tieniu Tan, Liang Wang, and Steve Maybank. 2004. A survey on visual surveillance of object motion and behaviors. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 34, 3 (2004), 334–352.
- [26] Jane Jacobs. 1961. *The death and life of great American cities*. Vintage.
- [27] Aditya Jitta and Arto Klami. 2017. Partially hidden Markov models for privacy-preserving modeling of indoor trajectories. *Neurocomputing* 266, Supplement C (2017), 196 – 205.
- [28] Jon Kerridge, Julian Hine, and Marcus Wigan. 2001. Agent-based modelling of pedestrian movements: the questions that need to be asked and answered. *Environment and planning B: Planning and design* 28, 3 (2001), 327–341.
- [29] Kris M Kitani, Brian D Ziebart, James Andrew Bagnell, and Martial Hebert. 2012. Activity forecasting. In *European Conference on Computer Vision*. Springer, 201–214.
- [30] Tian Lan, Tsung-Chuan Chen, and Silvio Savarese. 2014. A hierarchical representation for future action prediction. In *European Conference on Computer Vision*. Springer, 689–704.
- [31] Nicholas D. Lane, Emiliano Miluzzo, Hong Lu, Daniel Peebles, Tanzeem Choudhury, and Andrew T. Campbell. 2010. A Survey of Mobile Phone Sensing. *IEEE Communications Magazine* 48, 9 (2010), 140 –150.
- [32] Jeffrey S. Larson, Eric. T. Bradlow, and Peter S. Fader. 2005. An exploratory look at supermarket shopping paths. *International Journal of Research in Marketing* 22 (2005), 395–414.
- [33] David Lazer, Ryan Kennedy, Gary King, and Alessandro Vespignani. 2014. The Parable of Google Flu: Traps in Big Data Analysis. *Science* 343, 14 March (2014), 1203–1205.
- [34] Sebastian O. H. Madgwick, Andrew J. L. Harrison, and Ravi Vaidyanathan. 2011. Estimation of IMU and MARG orientation using a gradient descent algorithm. In *Proceedings of the IEEE International Conference on Rehabilitation Robotics*. IEEE, 1–7.
- [35] FILIPA Matos Wunderlich. 2008. Walking and rhythmicity: Sensing urban space. *Journal of Urban Design* 13, 1 (2008), 125–139.
- [36] Vikas Mehta. 2007. Lively streets determining environmental characteristics to support social behavior. *Journal of planning education and research* 27, 2 (2007), 165–187.
- [37] Vikas Mehta. 2014. Evaluating Public Space. *Journal of Urban Design* 19, 1 (2014), 53 – 88.
- [38] M. Muaz and R. Mayrhofer. 2017. Smartphone-based Gait Recognition: From Authentication to Imitation. *IEEE Transactions on Mobile Computing* (2017).
- [39] Heeyun Park and Shigeru Satoh. 2006. Analysis on the Pedestrian Rambling Activities and Spatial Structure in Downtown. *Journal of Architecture and Planning (Transactions of AIJ)* 605 (2006), 143–150.
- [40] Placemeter. 2019 (accessed January 9, 2019). <https://www.placemeter.com/>. (2019 (accessed January 9, 2019)).
- [41] Quuppa. 2019 (accessed January 9, 2019). <http://quuppa.com/>. (2019 (accessed January 9, 2019)).
- [42] Anshul Rai, Krishna Kant, Venkata N. Padmanabhan, and Rijurekha Sen. 2012. Zee: zero-effort crowdsourcing for indoor localization. In *Proceedings of the 18th annual international conference on Mobile computing and networking*. ACM, 293–304.
- [43] Jonathan Reades, Francesco Calabrese, Andres Sevtsuk, and Carlo Ratti. 2007. Cellular census: Explorations in urban data collection. *IEEE Pervasive Computing* 6, 3 (2007), 30–38.
- [44] Michael S Ryoo. 2011. Human activity prediction: Early recognition of ongoing activities from streaming videos. In *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 1036–1043.
- [45] Mitsuru Sakamoto, Asuka Yamada, and Jun Ueno. 2005. 5423 Influence Factors and Swing Motion of Pedestrians Traffic Line at an Urban Pedestrian Space. *Summaries of technical papers of Annual Meeting Architectural Institute of Japan. E-1, Architectural planning and design I, Building types and community facilities, planning and design method building construction system human factor studies planning and design theory* 2005 (July 2005), 879–880. <https://ci.nii.ac.jp/naid/110007078361/en/>
- [46] Lorenz Schauer, Martin Werner, and Philipp Marcus. 2014. Estimating crowd densities and pedestrian flows using wi-fi and bluetooth. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous)*.

- [47] Thorsten Schelhorn, David O'Sullivan, Mordechai Haklay, and Mark Thurstain-Goodwin. 1999. STREETS: An agent-based pedestrian model. (1999).
- [48] Xiaowei Shao, Huijing Zhao, Katsuyuki Nakamura, Kyoichiro Katabira, Ryosuke Shibasaki, and Yuri Nakagawa. 2007. Detection and tracking of multiple pedestrians by using laser range scanners. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*.
- [49] Virginia P. Sisiopiku and D. Akin. 2003. Pedestrian behaviors at and perceptions towards various pedestrian facilities: An examination based on observation and survey data. *Transportation Research Part F: Traffic Psychology and Behaviour* 6 (2003), 249–274.
- [50] Kardi Teknomo, Yasushi Takeyama, and Hajime Inamura. 2000. Review on microscopic pedestrian simulation model. In *Proceedings Japan Society of Civil Engineering Conference*. 15–27.
- [51] Kevin Toohey and Matt Duckham. 2015. Trajectory similarity measures. *Sigspatial Special* 7, 1 (2015), 43–50.
- [52] Ubisense. 2019 (accessed January 9, 2019). <http://ubisense.net/en>. (2019 (accessed January 9, 2019)).
- [53] Walkbase. 2019 (accessed January 9, 2019). <http://www.walkbase.com/>. (2019 (accessed January 9, 2019)).
- [54] Jin-Feng Wang, A Stein, Bin-Bo Gao, and Yong Ge. 2012. A review of spatial sampling. *Spatial Statistics* 2 (2012), 1–14.
- [55] Limin Wang, Yu Qiao, and Xiaoou Tang. 2015. Action recognition with trajectory-pooled deep-convolutional descriptors. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4305–4314.
- [56] Jens Weppner and Paul Lukowicz. 2013. Bluetooth based collaborative crowd density estimation with mobile phones. In *Proceedings of the IEEE international conference on Pervasive computing and communications (PerCom)*.
- [57] William Hollingsworth Whyte. 1980. *The social life of small urban spaces*. Project for Public Spaces.
- [58] William Hollingsworth Whyte. 2000. *How to turn a place around*. Project for Public Spaces. 225 – 232 pages.
- [59] Fang-Jing Wu and Gurkan Solmaz. 2018. CrowdEstimator: Approximating Crowd Sizes with Multi-modal Data for Internet-of-Things Services. In *Proceedings of the 16th ACM International Conference on Mobile Systems, Applications and Services*.
- [60] Yoshihisa Yamashita, Kinya Yokawa, and Hisao Uchiyama. 2006. A Survey Based on Watching and Tracing Pedestrians at a Railway Terminal Station. *Transport Policy Studies Review* 9, 34 (2006).
- [61] Massimo Zancanaro, Tsvi Kuflik, Zvi Boger, Dina Goren-Bar, and Dan Goldwasser. 2007. Analyzing museum visitors' behavior patterns. In *International Conference on User Modeling*. Springer, 238–246.
- [62] Cong Zhang, Hongsheng Li, Xiaogang Wang, and Xiaokang Yang. 2015. Cross-scene crowd counting via deep convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 833–841.
- [63] Kathryn Zickuhr and Aaron Smith. 2012. Digital differences. *Pew Internet & American Life Project* 13 (2012).
- [64] Brian D Ziebart, Nathan Ratliff, Garratt Gallagher, Christoph Mertz, Kevin Peterson, J Andrew Bagnell, Martial Hebert, Anind K Dey, and Siddhartha Srinivasa. 2009. Planning-based prediction for pedestrians. In *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*. IEEE, 3931–3936.

Received xxx; revised xxx; accepted xxx