

# Wireless Signal Based Crowd Condition Estimation

Vom Fachbereich Informatik der  
Technischen Universität Kaiserslautern  
zur Verleihung des akademischen Grades  
Doktor der Ingenieurwissenschaften (Dr.-Ing.)

genehmigte Dissertation

von

**Jens Weppner**

Datum der wissenschaftlichen Aussprache: 18.01.2018

Dekan: Prof. Dr. Stefan Deßloch

Berichterstatter: Prof. Dr. Paul Lukowicz

Berichterstatter: Prof. Dr. rer. nat. Dr. h.c. mult. Wolfgang Wahlster





# Abstract

Crowd condition monitoring concerns the crowd safety and concerns business performance metrics. The research problem to be solved is a crowd condition estimation approach to enable and support the supervision of mass events by first-responders and marketing experts, but is also targeted towards supporting social scientists, journalists, historians, public relations experts, community leaders, and political researchers. Real-time insights of the crowd condition is desired for quick reactions and historic crowd conditions measurements are desired for profound post-event crowd condition analysis.

This thesis aims to provide a systematic understanding of different approaches for crowd condition estimation by relying on 2.4 GHz signals and its variation in crowds of people, proposes and categorizes possible sensing approaches, applies supervised machine learning algorithms, and demonstrates experimental evaluation results. We categorize four sensing approaches. Firstly, stationary sensors which are sensing crowd centric signals sources. Secondly, stationary sensors which are sensing other stationary signals sources (either opportunistic or special purpose signal sources). Thirdly, a few volunteers within the crowd equipped with sensors which are sensing other surrounding crowd centric device signals (either individually, in a single group or collaboratively) within a small region. Fourthly, a small subset of participants within the crowd equipped with sensors and roaming throughout a whole city to sense wireless crowd centric signals.

We present and evaluate an approach with meshed stationary sensors which were sensing crowd centric devices. This was demonstrated and empirically evaluated within an industrial project during three of the world-wide largest automotive exhibitions. With over 30 meshed stationary sensors in an optimized setup across  $6400\text{ m}^2$  we achieved a mean absolute error of the crowd density of just 0.0115

people per square meter which equals to an average of below 6% mean relative error from the ground truth. We validate the contextual crowd condition anomaly detection method during the visit of chancellor Mrs. Merkel and during a large press conference during the exhibition. We present the approach of opportunistically sensing stationary based wireless signal variations and validate this during the Hannover CeBIT exhibition with 80 opportunistic sources with a crowd condition estimation relative error of below 12% relying only on surrounding signals influenced by humans. Pursuing this approach we present an approach with dedicated signal sources and sensors to estimate the condition of shared office environments. We demonstrate methods being viable to even detect low density static crowds, such as people sitting at their desks, and evaluate this on an eight person office scenario. We present the approach of mobile crowd density estimation by a group of sensors detecting other crowd centric devices in the proximity with a classification accuracy of the crowd density of 66% (improvement of over 22% over a individual sensor) during the crowded Oktoberfest event. We propose a collaborative mobile sensing approach which makes the system more robust against variations that may result from the background of the people rather than the crowd condition with differential features taking information about the link structure between actively scanning devices, the ratio between values observed by different devices, ratio of discovered crowd devices over time, teamwise diversity of discovered devices, number of semi-continuous device visibility periods, and device visibility durations into account. We validate the approach on multiple experiments including the Kaiserslautern European soccer championship public viewing event and evaluated the collaborative mobile sensing approach with a crowd condition estimation accuracy of 77% while outperforming previous methods by 21%. We present the feasibility of deploying the wireless crowd condition sensing approach to a citywide scale during an event in Zurich with 971 actively sensing participants and outperformed the reference method by 24% in average.

Table 1 presents the full list of performed crowd condition estimation experiments discussed within this thesis and beyond.

TABLE 1 Performed crowd condition estimation experiments.

---

2010	• Munich Oktoberfest	9 participants, 3 days
2010	• Munich Allianz arena soccer match fan arrivals	6 participants, 2 days
2010	• Passau Dult fair	12 participants, 3 days
2010	• Malta Farsons festival	10 participants, 3 days
2010	• London Wembley Stadium fan arrivals	6 participants, 1 day
2011	• Passau shopping center	2 participants, 1 day
2012	• Public viewing Kaiserslautern	20 participants, 2 days
2012	• Kaiserslautern fair	20 participants, 2 days
2012	• Kaiserslautern 1.FCK soccer match fan arrivals	5 participants, 2 days
2013	• Kaiserslautern Altstadtfest	3 stationary scanners, 3 days
2013	• Kaiserslautern 1.FCK soccer match fan arrivals	12 participants, 1 day
2013	• Zurich Zuerifaescht	971 participants, 3 days
2014	• Dresden Volkswagen V-Day event	250 participants, 1 day
2014	• Hannover CeBIT DFKI	14 stationary scanners, 5 days
2015	• Geneva Volkswagen Auto Salon	15 stationary scanners, 13 days
2015	• Shanghai Volkswagen Auto Show	20 stationary scanners, 20 days
2015	• Frankfurt Volkswagen IAA	33 stationary scanners, 13 days
2016	• DFKI shared office spaces BOSCH	25 participants, 25 days



# Acknowledgments

My time at University of Passau and at the German Research Center of Artificial Intelligence (DFKI) has been an inspiring and rewarding experience. I appreciate Prof. Dr. Paul Lukowicz, my academic supervisor, for giving me the opportunity to join his research group, offering his support, trust and freedom in my work. I thank him for his valuable research input, guidance and excellent inspiration over the past years as well as for the possibility of working in manifold outstanding research projects. I could always rely on his enormous experience and his advice. I would also like to thank Prof. Dr. rer. nat. Dr. h.c. mult. Wolfgang Wahlster for co-examining my PhD thesis.

Outside the lab I particularly thank Dr. Oliver Brdiczka and Dr. Kurt Partridge at the Palo Alto Research Center (PARC, California, USA) with whom I had many fruitful discussions and created my first patent during the time of my research internship.

I want to thank all members of the lab for their group spirit and the supportive environment and inspiring research culture. Especially, I would like to thank Orkhan Amiraslanov, Gernot Bahle, Dr. David Bannach, Dr. Gerald Bauer, Benjamin Bischke, Prof. Dr. Jingyuan Cheng, Vitor Fortes, Tobias Franke, Agnes Grünerbl, Peter Hevesi, Marco Hirsch, Dr. George Kampis, Dr. Kamil Kloch, Matthias Kreil, Prof. Dr. Kai Kunze, Christian Olczak, Dr. Gerald Pirkl, Andreas Poxrucker, and Bo Zhou for being great colleagues.

I also want to thank the countless students and participants of the experiments. Most importantly, I want to thank the most significant people to me, my family and my fiancée Elke Willy. They have always been there, supporting me in every way possible and made this endeavor possible.

Jens Weppner

Kaiserslautern, 2017



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivation . . . . .	2
1.2	Aims of the Thesis . . . . .	3
1.3	Related Work . . . . .	4
1.3.1	Crowd Condition . . . . .	5
1.3.2	Crowd Condition Sensing Technologies . . . . .	6
1.4	Thesis Outline . . . . .	8
1.5	Selected List of Publications . . . . .	12
<b>2</b>	<b>Crowd Condition Estimation with Stationary Scanners Opportunistically Scanning Crowd Devices</b>	<b>13</b>
2.1	Introduction . . . . .	14
2.2	Problem Statement . . . . .	15
2.3	Chapter Overview . . . . .	16
2.4	Related Work . . . . .	17
2.5	Opportunistic Scanning of Crowd Devices . . . . .	21
2.5.1	Opportunistic Bluetooth Discovery . . . . .	22
2.5.2	Opportunistic WiFi Probe Receiving . . . . .	23
2.6	Experiment Context . . . . .	25
2.6.1	Experiment Locations . . . . .	26
2.6.2	Data Set . . . . .	28
2.7	Approach . . . . .	32
2.7.1	Embedded Scanner Unit . . . . .	32
2.7.2	Meshed Scanner Setup . . . . .	34
2.8	Methods and Results . . . . .	36



*Contents*

2.8.1	Ground Truth . . . . .	37
2.8.2	Methods of Aggregated Crowd Density Estimation . . . . .	38
2.8.3	Aggregated Crowd Density Estimation Results . . . . .	45
2.8.4	Methods and Results of Local Crowd Density Estimation . . . . .	50
2.8.5	Methods and Results of Local Crowd Movement Estimation . . . . .	54
2.8.6	Methods and Results of Contextual Crowd Condition Anomaly Detection . . . . .	59
2.9	Conclusions . . . . .	64
<b>3</b>	<b>Crowd Condition Estimation with Stationary Sensors Scanning Stationary Devices</b>	<b>65</b>
3.1	Problem Statement . . . . .	66
3.2	Chapter Overview . . . . .	66
3.3	Related Work . . . . .	67
3.4	Signal Path Loss and Signal Propagation . . . . .	68
3.5	Opportunistic Scanning of Ambient Static Devices . . . . .	70
3.5.1	CeBIT Experiment Context and Data Set . . . . .	70
3.5.2	Methods and Results . . . . .	71
3.6	Scanning of Special Purpose Devices . . . . .	76
3.6.1	DFKI Experimental Environment and Hardware . . . . .	76
3.6.2	Methods . . . . .	78
3.6.3	Experimental Validation and Aims . . . . .	80
3.7	Conclusions . . . . .	88
<b>4</b>	<b>Crowd Condition Estimation with Mobile Scanners Opportunis- tically Scanning Crowd Devices</b>	<b>89</b>
4.1	Introduction . . . . .	90
4.2	Chapter Overview and Contributions . . . . .	91
4.3	Related Work . . . . .	93
4.4	Approach . . . . .	96
4.5	General Considerations . . . . .	99
4.6	Oktoberfest Experiment . . . . .	100
4.6.1	Individual Scanner Feature Vector . . . . .	100

4.6.2	Group Formation Scanning Feature Vector . . . . .	101
4.6.3	Machine Learning . . . . .	104
4.6.4	Experimental Environment and Data Collection Process . .	105
4.6.5	Data Distribution . . . . .	106
4.6.6	Experimental Validation Results . . . . .	108
4.7	Collaborative Scanning - Public Viewing Experiment . . . . .	113
4.7.1	Experimental Environment and Experimental Setup . . . .	114
4.7.2	Feature Vector . . . . .	117
4.7.3	Experimental Validation . . . . .	127
4.7.4	Results . . . . .	128
4.8	Conclusion . . . . .	133
<b>5</b>	<b>Participatory Citywide Sensing</b>	<b>135</b>
5.1	Introduction . . . . .	136
5.2	Motivation and Problem Statement . . . . .	139
5.3	Chapter Overview and Contributions . . . . .	140
5.4	Related Work . . . . .	141
5.4.1	Participatory Sensing . . . . .	141
5.4.2	Crowd Monitoring . . . . .	141
5.4.3	Bluetooth Scanning . . . . .	142
5.5	Data Set . . . . .	143
5.5.1	Experiment Advertising Campaign and Distribution . . . .	144
5.5.2	Privacy Policy and Anonymization Approach . . . . .	145
5.5.3	Experiment Procedure and Data Collection Process . . . .	146
5.5.4	Data Characteristics . . . . .	147
5.6	Citywide Area based Crowd Density Estimation . . . . .	151
5.6.1	General Principle . . . . .	151
5.6.2	Advanced Method . . . . .	154
5.6.3	Evaluation and Results . . . . .	157
5.7	Crowd Motion Characteristics . . . . .	161
5.8	Citywide Grid-based Crowd Density Estimation . . . . .	165
5.8.1	General Considerations . . . . .	166
5.8.2	Definition of Crowd Class and Ground Truth . . . . .	168

*Contents*

5.8.3	Methods and Machine Learning . . . . .	169
5.8.4	Evaluation Results . . . . .	175
5.9	Conclusion . . . . .	184
	<b>List of Figures</b>	<b>188</b>
	<b>List of Tables</b>	<b>190</b>
	<b>Bibliography</b>	<b>204</b>
	<b>Curriculum Vitae</b>	

# 1

## Introduction

## 1.1 Motivation

With approximately 21 stampedes in the last 5 years and 3000 attributed deaths reported globally, high crowd densities are a severe danger if not recognized immediately. The theory behind crowd stampedes and evolving crowd conditions is studied in a related research area where among others the critical crowd density thresholding values are determined with simulation models. However, an important component towards mastering crowd conditions is obtaining the actual values of the crowd condition in reality. The main motivation for this thesis is the importance of automatically estimating crowd conditions, as during the ‘Loveparade’ event in Duisburg in 2010 a fatal catastrophe happened due to crowd conditions the supervisory authorities were not fully aware of. Complications are that manual observations are difficult and common image recognition issues exist for observing the crowd condition. In the ubiquitous computing age we are surrounded by wireless signals - especially in the industrial, scientific and medical (ISM) radio band of 2.4 GHz. The ubiquitous computing revolution has been inspecting wireless signals for social inferring. Wireless signals have attributes which are likewise advantageous and disadvantageous as the human body absorbs signals in the 2.4 GHz spectrum. Before, sensing of wireless signals was used in social sciences for social inferring where the research community aimed towards detecting people in the proximity. However, the complications for detailed crowd condition estimation are the signal attenuation, the issue of sufficient statistics and the cultural factors. At 2.4 GHz the human body (made of about 70% water) has a high absorption coefficient. This means that in a dense crowd (where we would expect to have good statistics) the effective scan range is reduced and therefore ‘falsifying’ the results. As a result the range of an RF signal is not constant as the signals are attenuated by (a) free-space and by (b) objects. In an ideal constant environment the signal propagation loss  $L_p$  is given by  $L_p = 20 \log_{10}(f) + 10 n \log_{10}(d) 10^{-3} - 27.55 [dB]$  where  $f$  is the transmission frequency,  $n$  is the path loss exponent depending on the obstruction and  $d$  is the distance in meters. Without having a complete model of the dynamic environment we cannot solve the equation to get the distance value and vice versa we cannot solve the equation to get the path loss exponent to know how many people are in-between a signal link. Therefore, empirical evaluations at different

crowd densities are necessary, which this thesis proposes.

## 1.2 Aims of the Thesis

So far, Bluetooth and WiFi crowd condition research is fragmented, with most work focusing on individual narrow problems, or does not consider cultural effects, or the presented concepts were not evaluated with ground truth. This thesis aims to change that by providing a systematic understanding both in a theoretical and a practical way and presents the feasibility, describes the experimental setup, describes a comprehensive set of methods to deal with different challenges, and illustrates how such methods are evaluated.

While this thesis focuses on novel wireless signal based crowd condition methods and its evaluation we neither do provide a classical exhaustive survey nor do we provide a multidisciplinary review. This thesis aims at analyzing signals from common mobile devices equipped with wireless 2.4 GHz interfaces including the Bluetooth protocol and WiFi protocol. There is a lot of current interest in this topic but not much work in the field, yet. Despite the growing research interest in this area, many problems remain unsolved. This is particularly true for the case apart from specific scenarios such as queues. The first problem this thesis deals with is how to estimate the crowd condition from a continuous sequence of signal measurements. The second problem, once we have a solution to the estimation problem, is how to tell how good this solution is - how do we measure its performance. The goal is to help researchers make their wireless signal based crowd condition estimation systems more robust with respect to absorption and cultural effects and to provide evaluations serving as reference for future methods. The main questions to be solved include:

- How to deal with opportunistic passively received signals (from peoples' mobiles) not being received at deterministically time periods? How to setup a meshed network of stationary sensors and extract information from the intermittent individual signals but large mass of total signals?
- How to deal with potential scenarios where no signals from peoples' mobiles

exist? How can other opportunistic ambient signals or special purpose signals be used for estimating the crowd condition based on passive signal alterations?

- How to make the system more robust against variations in the number of discoverable devices that may result from the background of the people in the crowd rather than the crowd density?
- How to not just rely on the number of devices seen by a scan, but also take into account information about average observed signal strength and the variance in both the signal strength and the number of devices?
- How to benefit of combining the information from several mobile sensors/devices carried by different close by users, rather than on an individual scanner?
- How to combine the collaborative sensor information from several mobile phones carried by different groups of static and dynamic intermittently close by users (only below 0.2 % of all people are equipped with a Bluetooth scanning mobile phone)?
- How do the proposed methods of extracting information about the link structure between actively scanning Bluetooth devices, ratio of discovered devices in the current scan window to previous scan windows, teamwise diversity of discovered devices, number of semi-continuous device visibility periods, and device visibility durations perform in crowd density estimation?
- How does such an approach with mobile sensors perform in unconstrained city scale environments where participants are not students following well defined motion patterns but ‘normal people going about their business’? What does it take to improve the system performance under such conditions? Is it possible to go beyond density estimation towards the recognition of motion patterns even through the owners of the discovered devices (who do not actively participate in the data collection and do not provide GPS data) from whose we have the approximate location but have no motion information?

### 1.3 Related Work

A lot of related work has been done in the large field of defining, estimating, processing, and interpreting the crowd condition. In this related work section we

present work in the order from general to specific including fundamental work, general mobile sensing, and crowd condition sensing technologies. Wireless sensing has its roots in social relationships [10], to track people within a city [71, 1], to estimate waiting times in queues [97], to uncover complex social systems [31], to analyze people behavior [57], and mobile crowd sensing [75, 100]. Since the approaches within this thesis are manifold, specific wireless signal based crowd condition estimation related work is presented along the chapters of this thesis.

### **1.3.1 Crowd Condition**

The crowd condition can be described by various characteristics including the density of a crowd which is defined as the number of people in relation to a unit of space. Another secondary characteristic is the crowd movement which indicates either which fraction of the crowd is moving or how fast the crowd is moving in average. Besides the spatial crowd condition, also the temporal changes (increasing or decreasing crowd density by arriving or leaving people) are an indicator of the crowd condition which can also serve to detect temporarily contextual crowd anomalies. Related work has been investigating human stampedes such as during the Love Parade in 2010 in Duisburg [55]. Krausz and Bauckhage have shown that in areas of extremely high crowd density the movement of a person is affecting other people in proximity and as a result of the high crowd density shock waves occur and propagate through the crowd. When people are moved by the crowd and cannot control their own motion anymore people lose their balance and typically fall down in a shock wave to get crushed and suffocate [55]. In related work the need for a precise crowd density estimation to correctly assess the dangerousness of a situation is clarified [74].

In a completely different branch of the business analytics, the crowd condition can be an intelligence gathering tool to provide valuable indication about the interest of customers through quantifying the number of individuals browsing a product, or the customers during different times of the day [90]. The information gathered can then be used to optimize the need for staff, the floor plan layout, and product display optimization.



### 1.3.2 Crowd Condition Sensing Technologies

**Manual Crowd Counting** An early proposed simplistic method to manually estimate the crowd density was introduced in 1967 by Jacobs [46]. He was a journalism lecturer at the University of California at Berkeley and viewed many student demonstrations from his office above a plaza. He manually obtained the number of people within a square of stone pavement lines of around 22 foot (6.7 meters) and then he converted the value to the people per square foot unit. Today this simplistic method serves as the foundation for manually extracting the ground truth to be used for machine learning training and validation. Nowadays, this approach often performed by humans reviewing CCTV footage and counting people in a defined area and extract the crowd density [67, 80]. A different manual approach is counting people with digital clickers [32] at entrance/exit gates and infer the crowd density within the closed area.

**Computer Vision** A long history of the computer vision approach of crowd condition estimation work exists. The images or video streams are most often sourced from CCTV footage or aerial images. The approach includes visual analysis by itself or includes a combination of visual and thermal cameras [4, 96]. Two principle computer vision approaches exist: the direct approach (i.e. human detection) and the indirect approach (i.e. pixel-based, texture-based, and corner points based analysis). Direct approaches include methods based on the detection of fully visible humans [82, 48] or the head/face detection [60, 42, 21]. Indirect approaches include methods based on foreground pixels counting after background image subtraction [27, 64], [47], methods based on texture features analysis [66], methods based on histograms of edge orientations [20], or methods based on moving corner points to estimate the number of moving people [3]. A multitude of visual crowd flow tracking methods exist such as a frame difference algorithm [59] or optical flow methods [6].

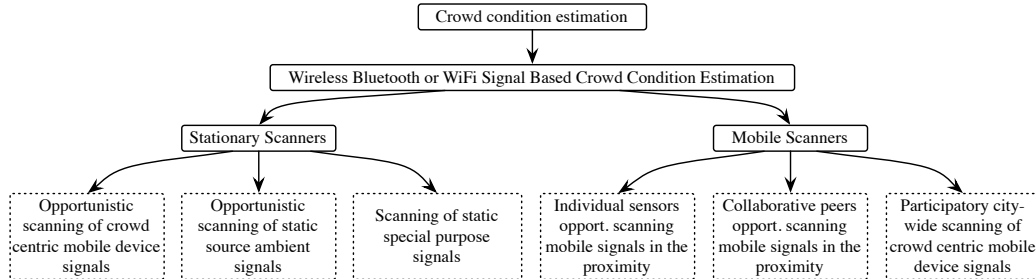
Despite the recent advances, along many computer vision approaches, it is referred to challenges with perspective problems, occlusions, and restricted light conditions. Specifically, edge-based features can be highly incorrect in the presence of complicated background and uneven textures of human clothes. The foreground and

background segmentation process becomes a more difficult task in crowded scenes. Also, extracting features from video streams is referred to be very time consuming for a subset of computer vision approaches.

**Other Stationary Sensors** Related work on crowd condition estimation also involves other technologies. The application range of technologies in related work is scattered. Two general approaches exist: automatic counting of people at entrance/exit gates and inferring the crowd density or estimating the crowd density directly. Methods include infrared barriers at gates [32, 45], personal radio-frequency identification (RFID) card detection systems [69, 84, 28], passive infrared (PIR) sensors for movement detection [94], pressure sensors [22], carbon dioxide sensors [63, 7], audio sensors [49], floor pressure sensors with footprint detection [69], seismic sensors [26], Kinect sensor [24], or radar sensors [23].

**Other Mobile Sensors** Since the rise of smartphones, general mobile crowd sensing is a vividly studied field in ubiquitous computing. Related work on mobile phone sensing in general is far reaching. Crowd sensing includes multiple intentions besides crowd condition estimation. Related work in the area of mobile phone sensing includes mechanisms to recruit crowd sensing participants [112]. Such mechanisms are primarily based on incentives towards collecting sensor data. Multiple mobile phone sensing concepts exist [56] including different potential mobile phone sensors, different sensing scales (individual, group, community), application distribution, sensing paradigms (either participatory sensing and being actively involved to opportunistic sensing) and mobile phone sensing architectures. A set of applications exists which uses smartphones as sensors for environment monitoring, traffic monitoring, human mobility behavior, interesting location discovery, public health, and social interaction detection [51]. In related work, mobile crowd condition estimation is accomplished with active user participation by [11, 105, 106] while relying just on GPS locations collected by a smartphone application and sending the data to a server. However, the challenge lies in the difficulty of recruiting a large number of participants to fulfill a sample large enough to represent the crowd condition.

## 1.4 Thesis Outline



**Figure 1.1.** The full thesis ontology of the considered wireless signal based crowd condition estimation methods.

The thesis is building on the ontology shown in Figure 1.1. The thesis topics include meshed stationary sensors scanning for crowd based opportunistic devices, stationary sensors scanning for external signal interferences by people of signals opportunistic WiFi access points, stationary sensors scanning for external signal interferences by people of special purpose connection-oriented signal sources, mobile sensors in single group scanning or collaborative scanning of crowd opportunistic devices, and mobile participatory scanning in the citywide scale.

**Chapter 2** This chapter introduces meshed stationary sensors scanning for opportunistic mobile device signals from the crowd. We setup meshed stationary sensors and recorded three large scale, real life data sets from a car manufacturers exhibitions at Geneva (Switzerland), Shanghai (China), and Frankfurt (Germany). Providing nearly 90 million data points from a total of over 670 000 unique mobile devices. We systematically analyze the approach of monitoring crowd density and crowd flow in real world environments. Questions that we address include the mapping from the number of detected devices to the number of people. We have developed and evaluated methods and compare multiple machine learning methods for crowd density estimation and visualization that build on the insights from the analysis above. We describe methods for the contextual crowd condition anomaly detection validated on two

significant events during the experiment being a visit by the German chancellor Mrs. Merkel and a press conference. We achieve approximation error rates of just 0.0151 people per square meter at average crowd densities of 0.22 people per square meter and maximum crowd densities of 2.6 people per square meter (during the press conference) and maximum crowd densities of 0.45 people per square meter during public days. The real-time crowd condition estimation visualization application was deployed and monitored during the exhibitions.

**Chapter 3** This chapter presents the approach of analyzing signals by stationary scanners opportunistically sensing ambient signal sources with stationary sensors scanning for either a large number of opportunistic signal sources or special purpose signal sources. A method is proposed for estimating the number of people present within an exhibition hall by analyzing signal variations from over 80 ambient wireless access points. A large-scale experiment is presented for validating the methods. A fingerprinting method is proposed for estimating the number of people present in a room (i.e. in a shared office space). A connection-oriented special purpose signal source allows measurements beyond signal strength such as signal phase and signal-to-noise-ratio. The method relies on machine learning to estimate the precise number of people within shared office spaces.

**Chapter 4** This chapter describes the crowd condition estimation approach with mobile sensors opportunistically scanning crowd devices. Mobile sensors are carried by people in groups or collaboratively through different crowd densities. This is compared to the naive individual scanner approach of just counting crowd devices in the proximity. Relative features rely on signal attenuation based on people and promote crowd condition estimation independent of cultural factors. This chapter includes evaluations based on multiple extensive data sets. We present a Bluetooth scan based method that can detect different discrete crowd densities. The main contributions beyond the above related work are to not just rely on the number of devices seen by a scan, but also take

into account information about the average observed signal strength and the variance in both the signal strength and the number of devices. This makes the system more robust against variations in the number of discoverable devices that may result from the background of the people in the crowd rather than the crowd density. We investigate the benefit of combining the information from several devices carried by different close by users, rather than on an individual scanner. We evaluate the method on a data set recorded during three days at the famous Munich Oktoberfest festival which is attended by hundreds of thousands of visitors from all over the world per day. We introduce new collaborative concepts of multiple teams walking intermittently nearby and scanning each other in addition to the previous method. The main contributions beyond the related work include also information about the link structure between actively scanning Bluetooth devices, ratio of discovered devices in the current scan window to previous scan windows, teamwise diversity of discovered devices, number of semi-continuous device visibility periods, and device visibility durations. We propose a method to combine the collaborative sensor information from several mobile phones carried by different groups of static and dynamic intermittently close by users (only 0.2% of all people are equipped with a Bluetooth scanning mobile phone) to determine the crowd density in an area of  $2500 m^2$ . We evaluate the collaborative method on a data set recorded during three days at the European soccer championship public viewing event in Kaiserslautern which is attended by thousands of visitors. Looking at seven discrete densities that cover the range from a nearly empty space (around 0.01 people per square meter) to dense crowd (above 2.0 people per square meter) we achieve recognition rates of over 75% using both relative and absolute features.

**Chapter 5** This chapter presents participatory wireless scanning which scales city-wide. Participative scanning follows a self-organizing citywide scanning approach. Participants are not students following well defined motion patterns but normal people going about their business. We evaluate the

crowd density estimation and crowd flow estimation methods on a data set consisting of nearly 200 000 discoveries from nearly 1000 scanning devices recorded during a three day citywide festival in Zurich. The data set also includes ground truth of 23 million GPS location points from nearly 30 000 users. The data set was used to compare the naive crowd density estimation method (extrapolating from the number of seen devices) with a more advanced method that goes beyond absolute numbers towards relative features that are more robust against statistical variations of the number of devices present at a given density. We analyzed the crowd condition estimation within 12 pre-defined areas during the Zurich event. A typical area has the size of thousand to tens of thousands of square meters. Areas are defined manually as being thematically associated. We also present an approach for the citywide grid-based area-independent crowd density estimation, visualizations, and validations.

## 1.5 Selected List of Publications

The selected list of publications, which this thesis is building on, is presented in Table 1.1.

**Table 1.1.** Selected List of Publications.

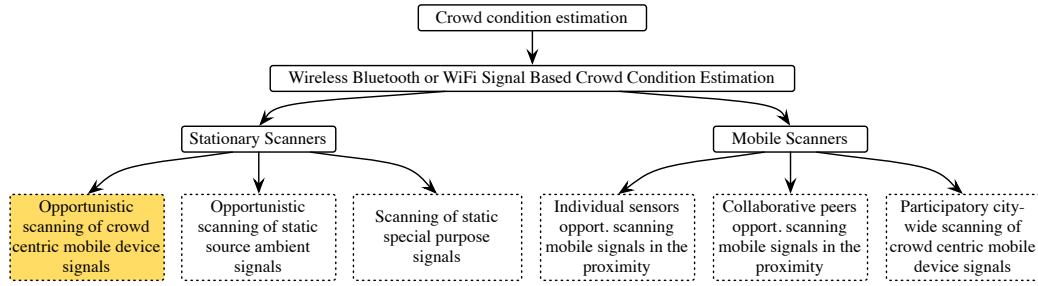
Chapter	Publication
2	Jens Weppner, Benjamin Bischke, and Paul Lukowicz. Monitoring crowd condition in public spaces by tracking mobile consumer devices with wifi interface. In <i>Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct. International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp-16), September 12-16, Heidelberg, Germany, UbiComp '16</i> , pages 1363–1371. ACM, 2016
3	Jens Weppner, Benjamin Bischke, and Paul Lukowicz. Sensing room occupancy levels with ieee 802.11n wifi channel state information fingerprinting. <i>IEEE Sensors Letter</i> , 2017
4	Jens Weppner and Paul Lukowicz. Collaborative crowd density estimation with mobile phones. In <i>Second International Workshop on Sensing Applications on Mobile Phones. ACM Conference on Embedded Network Sensor Systems (SenSys-11), 9th, November 1, Seattle, USA</i> . Microsoft, ACM, 2011 Jens Weppner and Paul Lukowicz. Bluetooth based collaborative crowd density estimation with mobile phones. In <i>Proceedings of the Eleventh Annual IEEE International Conference on Pervasive Computing and Communications (Percom 2013)</i> , pages 193–200. IEEE, 3 2013
5	Jens Weppner, Paul Lukowicz, Ulf Blanke, and Gerhard Tröster. Participatory bluetooth scans serving as urban crowd probes. <i>Sensors Journal, IEEE</i> , 14(12):4196–4206, Dec 2014

# 2

## Crowd Condition Estimation with Stationary Scanners Opportunistically Scanning Crowd Devices

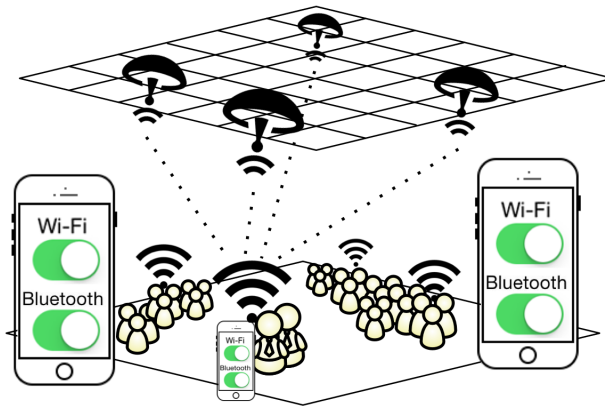
Jens Weppner, Benjamin Bischke, and Paul Lukowicz. Monitoring crowd condition in public spaces by tracking mobile consumer devices with wifi interface. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct. International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp-16), September 12-16, Heidelberg, Germany, UbiComp '16*, pages 1363–1371. ACM, 2016





**Figure 2.1.** Thesis outline and wireless signal based crowd condition estimation scanning ontology.

## 2.1 Introduction



**Figure 2.2.** Symbolic illustration of the meshed sensors mounted at the ceiling which passively scanned opportunistically wireless mobile crowd centric devices with enabled WiFi or Bluetooth interface.

The analysis of radio-frequency signals is a well known technique for human activity monitoring. In general, we distinguish three types of approaches (which may be used in isolation or in combination). First are systems where users’ mobile devices scan the environment for signals from stationary beacons such as for example WiFi access points or Bluetooth iBeacons. This is a basis for a whole range of indoor positioning systems (see related work). Second are systems where users’ mobile devices are used to detect the presence of other mobile devices. This approach

has been widely used especially for the tracing of social interactions. Thirdly, we have stationary scanners detecting, counting, and tracking mobile beacons carried by the users. Such mobile beacons can either be dedicated devices or the WiFi or Bluetooth interfaces of standard mobile devices such as smartphones or smartwatches. In this chapter we focus on the later. Specifically, we use carefully placed stationary WiFi and Bluetooth scanners with highly directional antennas to monitor crowd conditions during large scale public events (see symbolic setup in Figure 2.2). The advantage of the approach is that the crowd can be monitored without the need for user cooperation in the form of installing and starting an App or carrying a dedicated beacon. As outlined in the related work section below the general feasibility of the approach above has already been demonstrated in individual applications including some crowd density measurement (see related work). The contribution of this paper beyond such work is a systematic study and optimization of crowd monitoring methods using stationary scanners to track consumer devices with activated WiFi/Bluetooth interfaces on a large, real life data set that includes extensive video ground truth.

## **2.2 Problem Statement**

The research problem to be solved is a grid-based crowd condition monitoring method to enable and support the supervision of mass events by first-responders and by business intelligence personnel. Crowd condition monitoring campaigns concern the crowd safety and concerns marketing performance metrics. Real-time insights of the crowd condition is desired for quick reactions. Historic data evaluation is desired for profound crowd condition analysis. A reliable, repeatable and precise crowd condition measurement is required. The method needs to allow a continuous temporal monitoring and continuous spatial grid-based monitoring of the crowd condition. No active user participation should be required. The main condition for crowd condition estimation campaigns is a well defined event area to be equipped with stationary sensors. Important crowd condition measurements need to be supported in combination. The two fundamental crowd condition measurements are defined as the crowd density and the crowd movement state. A proposed crowd

condition measurement is estimated with an approximate value approaching the true value with the least possible estimation error.

To support the crowd condition surveillance, further insights beyond momentary measurements are needed. Intelligent processes are required to support the human embracing of circumstances requiring further action or giving historical insights. The temporal variations of the crowd condition measurements are another indicator and enables insight into the progression of the crowd condition. Indicators are needed to detect deviations in the crowd condition which could otherwise be overseen with momentary crowd monitoring. Deviations in the time series which are successfully recognized as significant crowd mutations are defined as crowd anomalies. Additional redundant measurements beyond crowd density and crowd movement need to be presented to support the reliability of detecting crowd anomalies.

## 2.3 Chapter Overview

This thesis chapter is organized as follows: We begin with presenting the existing related work on the specific field of opportunistic stationary scanning of crowd devices. Then the enabling foundations of wireless protocols for opportunistic scanning is shown. The experimental environments are declared to which the methods are applied to. The experimental environment description includes an overview about the scanner hardware, the scanner setup, the data sets, and the ground truth.

Next, the approaches follow including the definition of local and global crowd density estimation. The challenges in opportunistic scanning is explained and empirically demonstrated with multiple opportunistic localization methods under different conditions. We present a localization approximation being reliable under different conditions. Along the crowd density estimation the definition of crowd movement state estimation is presented.

In the following, the methods and results follow including the global crowd density estimation methods, applied machine learning principles, and present the quantitative machine learning regression results. We continue the methodological transfer

to local crowd density estimation, present result visualizations, and describe the qualitative results.

Next, the contextual crowd condition anomaly detection is presented with multiple redundant anomaly detection methods, and describe qualitative results.

1. We have recorded three large scale, real life data sets from car manufacturers exhibitions at Geneva (Switzerland), Shanghai (China), and Frankfurt (Germany). While the first two events were used for technological evaluation and visitor clustering the third event in Frankfurt during the IAA was directed towards crowd density estimation including video-based ground truth coverage. The data set is based on 33 directional scanners covering 9 ‘zones of interest’ and a total area of  $6400m^2$ . The scanners were running for 13 business days, providing nearly 111 million data points from a total of over 670 000 unique mobile devices. For 7 of the 13 days video ground truth has been recorded and extensively annotated.
2. We have used the data set to systematically analyze the limits and potential problems associated with monitoring crowd density and crowd flow in real world environments. Questions that we addressed include the mapping from the number of detected devices to the number of people (including the ability to generalize from a small number of ground truth points recorded on one day to other days), the ability to localize individuals in different conditions and the ability to reconstruct paths in different conditions.
3. We have developed and evaluated methods for crowd density estimation and visualization that build on the insights from the analysis above.

## 2.4 Related Work

Freudiger presented an experimental study of WiFi probe request [35]. He analyzed different smartphone brands and models on the burst rate of WiFi probes. They also analyzed the impact of a multitude of device states on the burst rate of WiFi probes. Such device states include the number of known WiFi SSIDs ever registered with the smartphone, the charging state, the screen on/off state, the state of the smartphone WiFi connection, and whether the user is currently in the WiFi settings

screen. Freudiger showed that the WiFi probe burst highly vary between a few seconds and 5 minutes with exceptional burst of up to 10 minutes. Freudiger also showed that 39% of the WiFi probes are missed (according to probe sequence numbers) which he assumes to be caused by the noisy nature of the wireless medium. However, the focus was put on the study of WiFi probes in general and not on the connection to the crowd condition.

Little and O'Brien presented within a CISCO White Paper the current possibilities of WiFi probe sensing [61]. They present use-cases for large sites including location analytics where they estimate the number of visitors, the amount of time they spend, and the frequency of their visits within the site. They also present advanced large site analytics which provide knowledge of movement patterns by these visitors while they are on the site. However, this White Paper does not present details of the methods and does not focus on crowd density estimation but rather on general crowded spots on large sites. A scientific evaluation with ground truth is not within the subject of the White Paper.

Li et al. demonstrated crowd condition estimation with just individual sensors in a small scale building environment [58]. The evaluation with ground truth was targeted whether the number of unique WiFi probe MAC falls below or exceeded the actual number of people. They analyzed the WiFi probe burst rate in dependence of the device vendor and device state (screen on, screen off, currently registered to a WiFi, not registered to a WiFi) and identified burst intervals between one second and individual maximum intervals of 20 minutes when a device is registered to a WiFi. They analyzed the impact of the human walking speed (slow, normal, jogging, and running) to the detection rate of WiFi probes. A decreasing detection rate with increasing speed was observed by individual sensors. This complies to the observation of potential long burst intervals which go beyond the range of individual sensors. They analyzed the trajectories between 6 individual scanners and discovered trajectory discovery rates between 28 % and 80 % for iOS devices depending on the device state. However, they did not focus on meshed stationary sensors and did not focus on the detailed ground truth based evaluation of the crowd density.

Handte et al. deployed a single WiFi probe sensor in a public transport bus for fill level estimation [38]. They were counting the people in a 30 minutes trip from

the start to the end of the bus route. After the trip they compared the measured number of unique WiFi probe devices with the ground truth. During the experiment the bus contained between 22 and 52 passengers and 20 % were visible. However, details about considered time windows are missing and the evaluation complexity is small.

Musa and Eriksson demonstrated WiFi probe based tracking of smartphones within a city over a distance of up to 2.8 km targeting road traffic congestion analysis and trajectory estimation [71]. They present a HMM approach estimating the location of a device between nodes and conclude that the localization accuracy depends on the geometry. However, they focussed on opportunistically detecting the location and did not focus on a detailed evaluation of the traffic density or crowd density. Multiple related work is focussing on the travel time extraction by detecting WiFi probe scans at two or more displaced sensors within a large scale citywide environment [1, 14, 87, 77]. Their approach is based on inferring the road congestion state based on the travel time. However, this approach has limited applicability towards crowd density estimation, as crowd can behave static and travel time extraction would be of limited benefit.

Other related work is focussing on detecting crowd with an individual sensor at a previously known specific location such as queues [97, 83]. For example, Schauer et al. evaluated the WiFi probe based queue waiting time by placing a WiFi probe sensor at the security check within the Munich airport. They used boarding pass scan numbers as ground truth for the number of people. The accuracy of both crowd density and pedestrian flow estimations was evaluated. A correlation of 0.75 is presented, however, the paper does not focus on an evaluation of a factor between ground truth and WiFi measurements.

Fukuzaki et al. collected WiFi probe requests from 20 sensors distributed widely in a shopping mall [36]. They used motion sensors at entrances for retrieving a calibration factor between 2.8 and 3.4 with an average error of 30 %. However, there is no mention of the expected accuracy of the automatic ground truth and how the motion sensor differentiated between inbound and outbound visitors for a correct count. The paper focussed on detecting single crowded spots within a medium-sized area and not on the precise crowd density with meshed sensors as presented in this thesis chapter.

Redondi et al. analyzed WiFi inter-probe periods for laptops and mobile devices. Their primary interest was to classify between mobiles and laptops by their WiFi probe pattern. They identified a broad range of inter-probe periods for continuously present devices on a university campus. They empirically extracted a probability of 95 % that mobile devices send a WiFi probe within 35 minutes and a probability of 80 % that mobile devices send a WiFi probe within 8 minutes [81]. However, it is not fully clear if devices have been in range all the time during the analysis and whether one can rely on this statistical information for further utilizations.

Zebra Technologies Corporation published a White Paper where they presented an analysis of WiFi probe MAC address randomization since iOS 8 [110] where they refer to the brief privacy enhancement statement with MAC randomization by Apple Inc. in [8]. They state that there are certain conditions under which the vendor may randomize, or not use, the device's static or real MAC address and in other conditions they may use the device's real MAC address. Zebra Technologies Corporation observed that the MAC randomization happens only in certain conditions that are not normal. They found that iPhone devices have this feature enabled and for random MAC to occur, the following conditions have to be met: 1. The phone wakes up from a sleep mode 2. The phone is not connected to WiFi 3. Cellular data is off 4. Location service is off. If all of the conditions are met the phone sends out its original MAC address initially for a few times and then sends out a probe request using random MAC addresses. In order to save battery life, an iPhone goes into sleep mode when all the services in the phone are inactive, not just when the screen is being switched off by the user. It is important to note that there are many applications that use location service, mail and message notifications which could keep the phone awake despite the screen being off. In addition, the cellular data is likely kept active and location services are likely to be kept active. So, in real life, it is very hard to make the phone go to sleep. Also, while the phone is connected to the WiFi network, the phone always uses the real MAC address in the Probe Requests. So, the MAC randomization is expected to have very little impact on analytics.

Stationary Bluetooth sensors were used in related work in specific situations. Kostakos et al. used Bluetooth discoveries to wirelessly detect and record end-to-end passenger journeys in public transport buses [53]. O'Neill et al. analyzed the

people throughput at entrance gates [76]. Versichele et al. presented a setup of stationary Bluetooth sensors during a city event where they observed Bluetooth device transitions between check-points and showed the temporal progression of the measurements [92]. Nicolai and Kenn investigated the discovery time of Bluetooth devices as well as the relation between number of people and number of discoverable Bluetooth devices [75]. They analyzed different locations and discovered a deviating people/device relation at different locations around the world. They used individual sensors and not a meshed sensor network and did not present the variations over multiple days. However, in general, related work lacks of a thorough experimental analysis with ground truth evaluation (except [75]) but focuses on the pure technical feasibility of Bluetooth discoveries.

In summary, while the general feasibility has been demonstrated before, this thesis chapter goes beyond previous research work and targets towards crowd density estimation and crowd movement estimation with meshed sensors with respect to a systematic study of various effects and comprehensive analysis of various crowd aspects in a large data set within a complex real world environment.

## 2.5 Opportunistic Scanning of Crowd Devices

We define wireless ‘scanning’ as the procedure to retrieve wireless device identifiers, independent of the wireless nature of WiFi or Bluetooth. The approach of opportunistic scanning of crowd devices is a virtual sensing technique (also called surrogate sensing) [62] used to provide a feasible and economical method to discover the crowd condition. The technique relies on measurements and learned parameters to estimate the desired quantity. The category of empirical virtual sensing relies on past measurements and the collateral ground truth observations. In general, empirical virtual sensing is based on regression techniques that can be implemented using a variety of machine learning modeling methods, such as linear regression, kernel regression, weighted least squares, support vector regression, or regression trees.

Opportunistic scanning (in the following synonymic for virtual sensing) is based on the assumption that a subset of the people within the crowd are equipped with a wireless enabled device. Nowadays, WiFi and Bluetooth equipped mobile



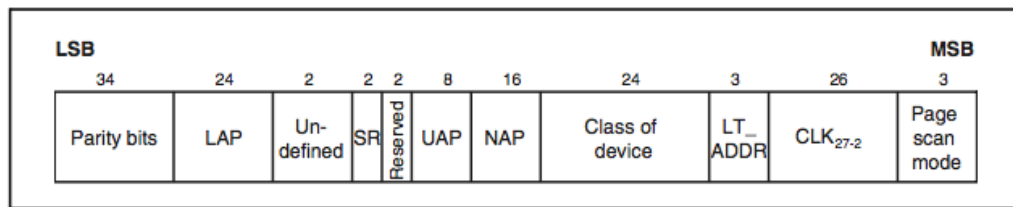
devices are ubiquitous. A large subset of visitors' devices have WiFi enabled. By experience some visitors disable WiFi permanently or temporally for energy saving. Our opportunistic scanning approach is built on top of 'Bluetooth' IEEE 802.15.1 standard and WiFi based on the IEEE 802.11 standard (currently Bluetooth SIG oversees the specification). Both protocols are coexisting on the 2.4 GHz frequency spectrum. Typical Bluetooth class 2 devices have a maximum permitted power of 2.5 mW and a signal range of up to 10 meters. Typical WiFi smartphones have a signal power of 7 mW [17] and a signal range of up to 50m–100 m. The standards were intended for wireless local area networking and for wireless communication between a mobile phone and a handsfree headset. For opportunistic scanning we take advantage of specific wireless protocols partitions of the specification. Although our implementations rely on high level application programming interfaces we define the protocols shortly.

### **2.5.1 Opportunistic Bluetooth Discovery**

For opportunistic Bluetooth scanning we built on top of the Bluetooth inquiry procedure [13]. The inquiry procedure was intended by the Bluetooth specification to discover a nearby device or even multiple devices. For Bluetooth this is typically the case for a mobile phone preparing to pair to a new Bluetooth headset. The inquiry procedure sends out an inquire, which is a request for nearby devices. Devices which are in discoverable mode (adjusted by the operating system or the user) issue an inquiry response. In total the inquiry procedure can take up to 10.24 seconds (128 train scans are repeated 4 times) after which all static nearby devices are deterministically known. Due to the frequency hopping architecture of the Bluetooth protocol the scanner (inquiring device) sends out a sequential inquiry on 32 different channels. The discoverable device in standby mode wakes up at least every 1.28 seconds and listens for at least 10 milliseconds. The listening device periodically listens on a single frequency for incoming inquiries and stays at this frequency long enough until the inquiring device has covered all frequencies. Devices always reply to each received inquiry with an inquire response. A repeated inquiry is repeatedly responded with an inquiry response. When a device receives an inquiry, it waits between 0 and 0.3 seconds before sending a Frequency Hopping

Synchronization packet in response. This is done to avoid a collision with another device that also wants to respond. The payload contains 144 bits plus 16 bits of CRC code. The information bits contain the UAP, LAP, and NAP. The unique identifier of the device is constructed by the LAP, UAP, and NAP as the 48-bit Bluetooth MAC address (see Figure 2.3). The UAP and NAP are assigned by the IEEE uniquely to each company. Each company assigns the LAP uniquely to each produced device.

With a specially equipped Bluetooth scanner unit we can continually disseminate Bluetooth inquiries and then process the inquiry responses.



**Figure 2.3.** Bluetooth inquiry response packet payload after active inquiry procedure. This packet is also called ‘frequency hopping synchronization packet’ to synchronize the frequency hopping scheme for the potential following communication. The relevant parts are the LAP and UAP. LAP: 24-bit field containing the lower address part of the responding device. UAP: 2-bit field containing the upper address part. NAP: 16-bit field containing the non-significant address part. Class of device: 24-bit field containing the device type. Source: ‘Specification of the Bluetooth protocol’ [13]

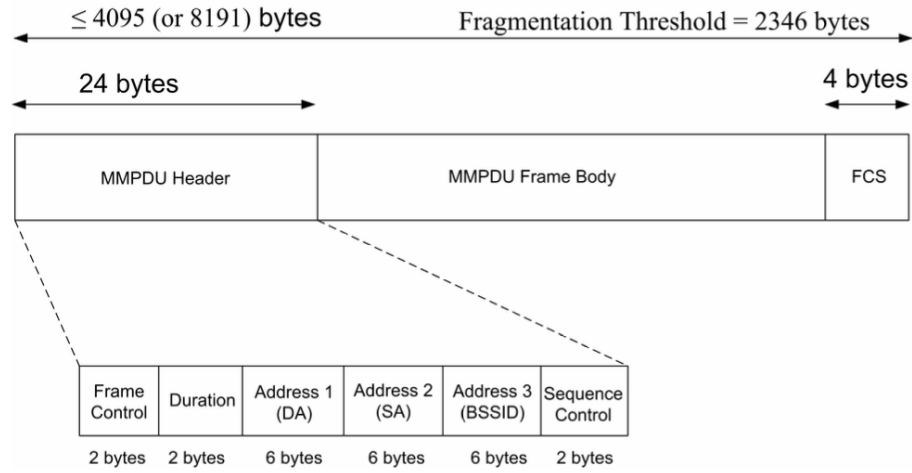
### 2.5.2 Opportunistic WiFi Probe Receiving

For opportunistic WiFi scanning we build on top of the WiFi scanning procedure [43]. Receiving WiFi devices is opportunistic in two ways: the devices in proximity are opportunistic and the receiving of device signals is opportunistic as the approach is fully connection-less. The concerning procedure was intended by the WiFi specification IEEE 802.11 to find a known network on arrival or to find a known network during roaming. For WiFi this is typically the case for a smartphone, tablet, embedded device, or laptop preparing to connect to a known access point periodically when arriving to a previously visited place. Even in

standby, mobile smartphones regularly wake up for trying to find a WiFi network for internet access in favor for a cellular connection. Two variants exist. ‘Passive scanning’ finds networks simply by listening for a ‘beacon frame’ from any access point (AP). This is not suitable for opportunistic scanning. Active scanning by the mobile device finds networks by sending a ‘probe request’ on each channel and waits for a ‘probe response’. The probe request is received by any access point, either known or unknown. AP probe response then announces its presence and Service Set Identifier (SSID also known as human readable WiFi name). The probe request is performed by a ‘MAC WiFi management frame’ also called ‘MAC Management Protocol Data Units’ (see MMPDU in Figure 2.4). The MMPDU header contains the MAC address (DA) of the wireless AP to receive this frame, the MAC address (SA) of the wireless host transmitting this frame, and the MAC address (BSSID) of the router interface which the AP is attached to [104]. The MMPDU body contains a requested SSID and supported data rates. The MMPDU body is not of further interest for opportunistic scanning.

As the WiFi scanner is purposely not a known WiFi device, no advancing connection to the scanner unit will take place by the host. Normal access points do not propagate received ‘probe request’ to the application layer because they are perceived as part of the WiFi management. The advantage about other connection oriented WiFi detections methods is the fact that the visitor does *not* need to interact by manually connecting to a specific AP. With a specially equipped WiFi scanner unit we can not just receive probe requests but process them further.

By extracting the non-significant the MAC address of the wireless host transmitting the ‘probe request’ from the MMPDU header the device vendor statistics are retrieved. In Table 2.2 we list the variety of device vendors discovered during the experiments.



**Figure 2.4.** The probe request is performed by a ‘MAC WiFi management frame’ (image source [104]) also called ‘MAC Management Protocol Data Units’. The MMPDU header contains the MAC address (DA) of the wireless AP to receive this frame, the MAC address (SA) of the wireless host transmitting this frame, and the MAC address (BSSID) of the router interface which the AP is attached to [104]. The MMPDU body contains a requested Service Set Identifier (SSID) and supported data rates.

## 2.6 Experiment Context

In this section we describe the experimental context including experiment locations, the meshed scanner setup, the scanner hardware and data processing pipeline, the data sets, and the collected ground truth used for qualitative visual evaluation and machine learning evaluation. We later refer to the experimental environment in section 2.7 and in section 2.8.

The experiments were performed in the context of a large industrial cooperation project funded by a large German car manufacturer with the aim of bringing new technology from academia to a large scale industrial-proof real-world campaign. The scientific exploitation of the experiments go far beyond the common scientific laboratory experiment scale. The experiments cover a large mass of visitors emitting WiFi probes and responding to Bluetooth discovery requests. We have a multitude of experiment days with different levels of the crowd condition. Different experiment days include diverse groups of people with a different background

and different mobile device configuration. The experimental environment includes simultaneous scanning of WiFi probes and Bluetooth inquiry responses. The experimental environment also includes ground truth with accurate manual human annotations extracted from the ground truth video stream. The experiments exceed previous experiments in the setup technology (meshed scanners), in scale (number of continuously observed devices), and in scientific evaluation complexity (ground truth evaluation with machine learning).

### 2.6.1 Experiment Locations

The industrial project included three experiment location across the world. The experiment locations were part of the largest automobile exhibitions in the world consisting of the exhibition in Geneva (Switzerland) titled ‘Geneva International Motor Show’, the exhibition in Shanghai (China) titled ‘Shanghai International Automobile Industry Exhibition’, and the exhibition in Frankfurt (Germany) titled ‘IAA’.

The Geneva motor show is an annual auto show held in March in the Swiss city of Geneva. Our experiment took place during the 13 days of the auto show in 2015. The exhibition is in the convention center ‘Geneva Palexpo’ located next to Geneva international airport. The exhibition is organized by the ‘Organisation Internationale des Constructeurs d’Automobiles’. The auto show is held since 1905 and hosted all major combustion engine models in the history of the automobile. Official visitor statistics describe between 34 984 and 87 192 unique visitors entering the exhibition during each day. Accumulated along 13 days a total of 683 681 visitors came to the exhibition.

The Shanghai auto show is an biennial exhibition held in April in the Chinese city of Shanghai. Our experiment took place during the 10 days of the auto show in 2015. The exhibition is placed in the convention center ‘National Center for Exhibition and Convention’ located next to the Shanghai Hongqiao international airport. The convention center consists of 8 large halls with two floors. The auto show is organized by a consortium of the Shanghai International Exhibition Co., Ltd., the China Association of Automobile Manufacturers, the China Council for Promotion of International Trade Shanghai Sub-Council, and the European co-organizer ‘Inter-

nationale Messe- und Ausstellungsdienste'. The auto show is held since 1985 and is the nation's oldest auto exhibition and being one of the top auto shows due to the expanding presence of foreign brands in the Chinese market. The exhibitions ranges from passenger cars, commercial vehicles, buses, trucks, new energy technology, automative parts, automative accessories, car maintenance products, auto supplies, manufacturing technology, to automotive computer systems. The exhibition space covers 350 000 square meters, includes 1185 exhibitors, 1343 vehicles on display, 109 world debut cars, 47 concept cars, and 103 new energy vehicles. Accumulated along 10 days a total of 928 000 visitors came to the exhibition.

The Frankfurt auto show is a biennial exhibition held in September in the German city of Frankfurt am Main. Our experiment took place during the 13 days of the auto show in 2015. The exhibition is placed in the convention center 'Messe Frankfurt ('Frankfurt Trade Fair') located in the heart of the financial and business centre Frankfurt. The auto show is organized by the 'Association of the German Automotive Industry (VDA)'. The auto show is held since 1897 and is the world's largest motor show. The convention center consists of 12 halls covering 367 000 square meters and additional free space of 96 000 square meters, includes 1103 exhibitors from 39 countries, 219 world premier cars. 11000 journalists from 106 countries attended the exhibition. Accumulated along 13 days a total of 931 700 visitors with an average age of 34 years came to the exhibition. The auto show was limited on three days to journalists, limited on two days to the professional audience and on nine days opened to general audience. The opening hours ranged from 09:00 to 19:00. The experiment sensing technology was setup within the large 6400 square meter booth of a large German car manufacturer within the half of the large exhibition hall 3. We set up the meshed scanners at each venue, while the Geneva venue and the Shanghai venues were used as technological prototypes and for visitor analytics which is out of the scope of this thesis, only the venue in Frankfurt was considered for the scientific crowd condition estimation with extensive ground truth information.

## 2.6.2 Data Set

During the experiments in Geneva, Shanghai, and Frankfurt we collected three extensive data sets containing up to 111 millions of scan entries per experiment. A single raw data entry in the data set contains of five elements: meshed network scanner identifier, timestamp, extracted MAC address vendor, received signal strength indicator, and the anonymized/hashed MAC.

We give an overview of the collected data sets by demonstrating extracted statistical information from and summarized them in Table 2.1. In the Geneva experiment we installed 15 Bluetooth/WiFi scanners and collected over 3.2 million Bluetooth inquiry responses, 16.2 thousand unique Bluetooth MAC addresses were detected at an average detection rate of 71 Bluetooth scans per minute. 8.1 million WiFi probes were received during the event, 113.7 thousand unique WiFi MAC addresses detected at an average detection rate of 282 per minute.

During the Shanghai experiment we installed 10 Bluetooth/WiFi scanners and collected over 1.2 million Bluetooth inquiry responses, 22.2 thousand unique Bluetooth MAC addresses were detected at an average detection rate of 185 per minute. 55.6 million WiFi probes were received during the event, 848.0 thousand unique WiFi MAC addresses detected at an average detection rate of 9312 WiFi scans per minute.

During the Frankfurt experiment we installed 33 Bluetooth/WiFi scanners and collected over 5.3 million Bluetooth inquiry responses, 31.2 thousand unique Bluetooth MAC addresses were detected at an average detection rate of 598 per minute. 111.0 million WiFi probes were received during the event, 670.0 thousand unique WiFi MAC addresses detected at an average detection rate of 9312 per minute.

We also gathered statistical device vendor information during the experiments and observed a highly diverging device vendors for WiFi probes and Bluetooth inquiry responses which we summarized in Table 2.2. The vendors Apple, Samsung, and Sony lead in the ranking and have a cumulated share of over 70% of WiFi enabled devices being detectable by outgoing WiFi probes. Other vendors follow in the WiFi probe ranking: HTC, Murata (manufacturing smartphone components for other vendors), Motorola, Microsoft, LG, Huawei, Nokia, Intel, RIM, Asustek, Hon Hai, and Azurewave. Other vendors are omitted. As noted before the vendor

identifier is extracted from the MAC address which can then be used for a lookup in IEEE MAC vendor tables. Vendors not being registered in the IEEE vendor list or randomized MAC addresses are not part of the statistics. The Bluetooth vendor ranking has a completely different order than the WiFi ranking. Nokia, RIM, and Samsung are the leaders with a share of over 80 % of Bluetooth enabled devices being detectable by Bluetooth inquiry responses. Other vendors follow in the Bluetooth ranking: Apple, Sony, LG, Microsoft, Huawei, Murata, Hin Hai, Intel, HTC, Motorola, Liteon, and Azurewave. Nevertheless, it is interesting to see that 6.83 % of Apple devices have Bluetooth enabled, as they are only in Bluetooth discoverable when the Bluetooth settings screen is active. These vendor statistics give a first hint that different groups of users are represented by Bluetooth and WiFi scans, as Bluetooth inquiry responses are mostly originating from traditional Nokia mobile phones and business centric BlackBerry RIM mobile phones.



**Table 2.1.** Experiment data set statistics of stationary opportunistic scanning of crowd devices. Upper table: Bluetooth data set statistics. Lower table: WiFi data set statistics. Table columns include the experiment duration, number of scanner units in the experiment setup, number of scans (including multiple scans of identical unique device), number of unique devices, mean scans per minute, and median scans per minute.

Event	Duration (days)	Meshed Scanners	Bluetooth scans/event	Bluetooth devices/event	Bluetooth scans/minute (mean)	Bluetooth scans/minute (median)
Hannover (DE) CeBIT	5	14	79 034	4599	94	89
Geneva (CH) Auto Salon	13	15	3 227 046	16 228	71	67
Shanghai (CN) Auto Show	10	20	1 211 637	22 280	185	196
Frankfurt (DE) IAA	13	33	5 396 517	31 278	598	562

Event	Duration (days)	Meshed Scanners	WiFi scans/event	WiFi devices/event	WiFi scans/minute (mean)	WiFi scans/minute (median)
Hannover (DE) CeBIT	5	14	1 845 217	48 949	485	485
Geneva (CH) Auto Salon	13	15	8 102 900	113 704	282	288
Shanghai (CN) Auto Show	10	20	55 624 490	848 051	9312	6519
Frankfurt (DE) IAA	13	33	111 031 490	670 610	12 974	10 365

**Table 2.2.** Scanned WiFi and Bluetooth crowd device vendors statistics.

<b>WiFi Device Vendor</b>	<b>Fraction</b>	<b>Bluetooth Device Vendor</b>	<b>Fraction</b>
Apple	44.59 %	Nokia	32.53 %
Samsung	22.98 %	RIM	30.20 %
Sony	6.19 %	Samsung	21.01 %
HTC	5.68 %	Apple	6.83 %
Murata	5.04 %	Sony	3.05 %
Motorola	4.75 %	LG	1.57 %
Microsoft	2.98 %	Microsoft	1.55 %
LG	2.96 %	Huawei	0.70 %
Huawei	1.73 %	Murata	0.63 %
Nokia	1.39 %	Hon Hai	0.59 %
Intel	0.70 %	Intel	0.46 %
RIM	0.48 %	HTC	0.46 %
Asustek	0.16 %	Motorola	0.18 %
Hon Hai	0.15 %	Liteon	0.17 %
Azurewave	0.08 %	Azurewave	0.03 %

## 2.7 Approach

In this section we present the fundamental approach towards our crowd condition estimation including the meshed scanner setup, the embedded scanner hardware, the definition of the global and grid-based crowd density estimation approach, the general crowd movement estimation approach, and localization approaches suitable for opportunistic wireless localization.

### 2.7.1 Embedded Scanner Unit

The embedded scanner unit consists of the hardware and core Bluetooth/WiFi scanning software to scan, process, and transfer scans to a central server. The embedded scanner hardware includes a compact, embedded, and passively cooled platform based on the Swiss PC Engines ALIX platform which is popular for special purpose network computing units. The mainboard has a small form factor of 15\*15 cm suitable to fit into a compact enclosures (see Figure 2.5a). The mainboard contains a 500 MHz AMD Geode CPU, 256 MB memory, CompactFlash card socket, miniPCI slots, USB connector, and Ethernet connector. WiFi and Bluetooth scan functionality is realized with miniPCI cards connected to the system with antenna cables connected to external enclosure connectors. The platform is running the ‘tinyBIOS’ open source system BIOS which is hosting a slimmed down embedded Debian Linux version. The embedded scanner unit is placed inside a robust enclosure allowing a deployment outside of the lab in real-world environments. A variety of external antenna directional antennas can be attached to the embedded unit. For the scenario in this thesis chapter we selected highly directional antennas with a theoretical wireless beam angle of 12 degrees. The embedded scanner software builds on top of open source software packages which are designed for wireless network security research and wireless network engineering. The core WiFi scan collection software is building on top of the open source ‘airodump’ software. The software contains a variety of functionalities including the functionality of presenting all wireless access points in range and the particular functionality of presenting received WiFi probe signals from wireless clients within the signal range. The intention of inspecting WiFi probes is manifold but not yet

introduced to large-scale meshed scanning targeting the crowd condition estimation. The details of WiFi probes are described in subsection 2.5.2. The functionality of the research tool ‘airodump’ goes beyond those functionalities and also offers packet capturing which we are not building on. Our software is querying the ‘airodump’ module in regular intervals to extract the currently received WiFi probe requests. For each received WiFi probe the following core scan processing is initiated:

- vendor lookup by MAC address from local IEEE vendor lookup table
- randomized WiFi MAC identification (relying on failed vendor lookup)
- time-slot based MAC hashing for anonymization (no real MAC address is stored)
- received signal strength indicator (RSSI) extraction
- creation of current timestamp
- creating set of information including the local scanner unit identifier (ascending number from 0 to 32), anonymized user identifier (hashed MAC), vendor, ‘randomized’ flag, RSSI, and timestamp
- transmitting set of information to central data base server (see Figure 2.6 on page 35)

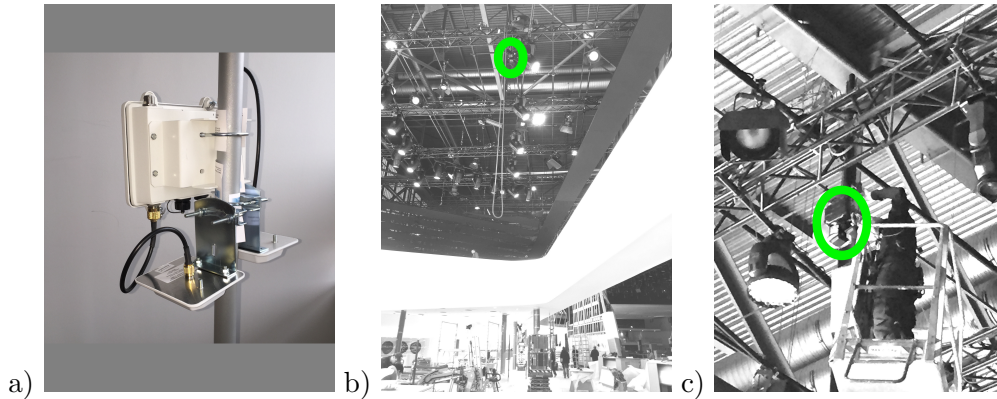
MAC addresses were checked for integrity by comparing the vendor identifier part of the MAC address to a IEEE vendor lookup table. When a valid vendor identifier existed it was marked as valid, otherwise marked as randomized/dynamic. For complying with legal authorities the MAC is anonymized by a hash function changing over time. After every 24 hours the hashing function changes for continuous anonymization. No real MAC address was persisted.

The core Bluetooth scan collection software is building on top of the open source ‘hcitool’ software. The software contains a variety of functionalities including the functionality for system administration, configuring Bluetooth connections, and sending special commands to Bluetooth devices. Special commands include displaying connected devices, inquiring/scanning remote devices, naming a remote device with a given Bluetooth address, printing supported features of the remote device, creating a baseband connection to the remote device, displaying received signal strength information for the connection to the device, requesting authentication for the device, and further lower-level functionalities. The Bluetooth scan software module is relying on the ‘hcitool’ and inquiring for Bluetooth devices. The inquiry

command is continually send to the ‘hcitool’. Together with the inquiry the received signal strength information command is executed. For each received Bluetooth inquiry response the following core scan processing is initiated:

- vendor lookup by MAC address from local IEEE vendor lookup table
- time-slot based Bluetooth MAC hashing for anonymization (no real MAC address is stored)
- received signal strength indicator (RSSI) extraction
- creation of current timestamp
- creating set of information including the local scanner unit identifier (ascending number from 0 to 32), anonymized user identifier (hashed MAC), vendor, RSSI, and timestamp
- transmitting set of information to central data base server (see Figure 2.6)

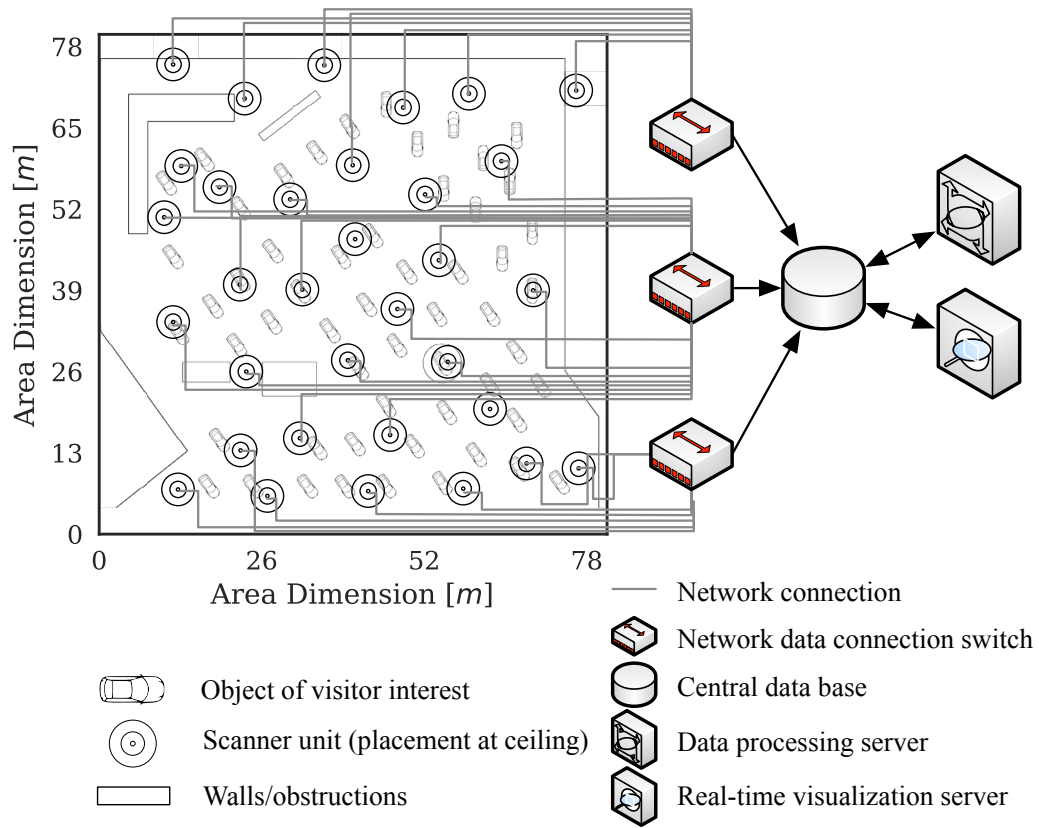
The main part of processing is performed by the backend servers accessing the raw central data base entries (see methods in section 2.8).



**Figure 2.5.** Photographs of a) the scanner unit in enclosure and connected external downward-directional WiFi/Bluetooth antennas, b) the installation preparations at the exhibition booth, c) the installation preparations in close-up.

## 2.7.2 Meshed Scanner Setup

Related work previously considered solitary WiFi probe scanners giving a hint at the device situation at a single spot. In contrast, this thesis chapter is building on



**Figure 2.6.** Stationary WiFi/Bluetooth scanner setup during the IAA (Frankfurt, Germany) for crowd density estimation including floor plan, meshed stationary scanner unit setup, objects of visitor interest, network infrastructure, central data base, data processing server, real-time visualization server, and object of visitor interest positions.

top of a meshed scanner setup suitable for crowd density and crowd flow estimation. With a meshed setup devices can be localized under certain conditions or assigned to a scanner cell. We later describe how the crowd density can be measured within such a scanner cell. We carefully planned the meshed scanner setup by considering the regions of visitor interest. Nine regions of interest existed each consisting of an exhibition element. We assigned one scanner specifically to each region of interest to avoid boundary effects of measurement fluctuations if two or more scanners are equidistant to a region of interest. The remaining 24 scanners were uniformly positioned across the exhibition area. In total 33 scanners were positioned with an

average distance of 14 meters and an average scanning zone of  $180\text{ m}^2$ . Figure 2.6 also visualizes the scanner distribution. One major issue in wireless transmissions in general is the influence of the signal by objects within the signal propagation path which is formed around the line of sight. When positioning a scanner on the ground significant shielding is caused by human bodies which can influence the localization within heterogenous crowd densities. We considered the approach of mounting the scanners on the ceiling. From the ceiling the scanners have an optimized free line of sight towards the device, and just constricted by the body of the inherent visitor and directly surrounding visitors. A secondary practical reason for mounting the scanners at the ceiling was the positioning freedom at the ceiling. Figure 2.5 shows a scanner unit and a scanner unit mounting preparations on the ceiling.

Each scanner was connected for maximum reliability via Ethernet cable over a cascade of network switches to a central data base server and pushed scan results to it (see Figure 2.6). The processing server and the visualization server accessed the central data base and did not need to communicate with each scanner individually.

## 2.8 Methods and Results

In this section we present the methods and results of the crowd density estimation and crowd movement estimation. The section includes the definition of the ground truth, the definition of aggregated and local crowd density estimation reference regions, the methods for aggregated crowd density estimation, the summary of the applied machine learning algorithms, and present the machine learning results of the aggregated crowd density estimation. We apply several different features (see methods in subsection 2.8.2) along with the extensive ground truth to several popular machine learning algorithms. We continue with transferring the methods to the local crowd density estimation, present crowd density visualizations, and describe the qualitative results on the local crowd density estimation. We continue with the transferring the methods to the local crowd movement estimation, present crowd movement level visualizations, and describe the qualitative results on the local crowd movement estimation. Finally, we present the anomaly detection

methods based on multiple crowd condition measurements and present qualitative results on two significant events during the experiment.

### **2.8.1 Ground Truth**

The notation of ‘ground truth’ is used in artificial intelligence machine learning in combination with supervised learning techniques. This serves as the ‘gold standard’ which is the best available information under reasonable conditions. Collecting ground truth refers to the process of gathering the proper objective data.

We collected ground truth during seven days of the experiment. We managed to temporally mount a single compact video camera (GoPro) while being supervised by a human. It was located at a height of 5 meters in direct vicinity of the area of interest. The view was diagonally directed towards the observed area. The camera was equipped with a wide angle lens, therefore covering almost the whole exhibition area. The video camera recorded video footage continuously and had to be re-charged regularly. Video data was extracted in regular intervals. Extracting the ground truth was a multi-step task. Firstly, frames were uniformly selected from the video with about 30 minutes to one hour of time difference. During observable high fluctuations of the crowd density the ground truth interval was reduced to 30 minutes. Secondly, each of these images were annotated in a labor-intensive task by students. For this task we relied on a graphical image processing tool as an annotation tool which allowed direct highlighting on a separate layer of the image. The annotations layer could then be exported for further processing. Popular specialized image annotation open source software tools failed because of frequent crashes when labeling over thousand entities within a single image. Annotations were defined as highlighting all visible people in the image. Up to 1200 annotations were performed per image. In total 42 444 annotations were made for in total 71 ground truth frames. Thirdly, annotation files were further processed, leaving the actual task of counting annotations to the software.



## 2.8.2 Methods of Aggregated Crowd Density Estimation

In this sub-section we present the required general data pre-processing steps, present the reference method serving as a baseline for later evaluation, present the time window based concept (important for later transfer to local crowd density estimation) with varying time ranges reflecting different potential visit durations, present the proposed WiFi based aggregated crowd density estimation methods, present the proposed Bluetooth based aggregated crowd density estimation methods, explain why these methods were used in distinction to other time series based analysis methods, and describe the machine learning training/validation process, and finally present the applied machine learning algorithms.

**Definition of Aggregated/Local Crowd Density Estimation** In general, the crowd density indicates the number of people within a unit area. The crowd density is i.e. used to describe the population density at different scales. A population density can be expressed as the number of inhabiting people per square kilometer. This can be measured within the boundary of a city or within a whole country and up to the world scale. Similarly, this thesis deals with crowd densities at different scales: 1. the crowd density within the whole observed area, and 2. the crowd density within parts of the observed area. We define the crowd density within the whole area (higher scale) as the ‘aggregated’ crowd density because it is originating from measurements of the whole meshed scanner network. We define the crowd density within parts (smaller scale) of the observed area as the ‘local’ crowd density. Commonly for such scales the unit of people per square meter is used. The aggregated crowd density scale is spanning over multiple scanners (in this experimental evaluation:  $6000m^2$ ) and represents the averaged number of people per square meter. The crowd density is reduced to a single value per observed meshed scanner area. This scale is useful when analyzing the evolving crowd density within a time based sequence within a simple line-plot. Within this thesis chapter the aggregated proposed crowd density methods are extensively evaluated with the supporting ground truth. The aggregated crowd density estimation methods are the foundation for the later localized crowd condition estimation methods which

are calibrated based on the former. The ‘local’ crowd density scale is referring to the area of one scanning zone and still represents the average number of people per square meter but within a much smaller scale of in average  $180m^2$ , which corresponds to a circular area with a radius of just 7.5 m or to a quadratic area width of just 13.5 m. This scale is useful when analyzing the distribution of the crowd density within a two-dimensional visualization, which is essential when analyzing which regions are more densely packed than others. Within this thesis chapter an aggregated crowd density estimation method is transferred to the local crowd density estimation, is visualized, and evaluated qualitatively within the capabilities of the supporting ground truth. The local crowd density estimation method is calibrated based on the results of the aggregated crowd density estimation.

**General Data Pre-Processing** Before the data is ready to be processed by the crowd density estimation methods, the data needed to be pre-processed including rejecting static devices and rejecting received signal from out of boundary. Signals can be received from mobile visitor devices and static devices. A static device is defined as not being attached to a person. Such static devices are embedded WiFi/Bluetooth enabled embedded components, WiFi/Bluetooth enabled demonstration tablets or WiFi/Bluetooth enabled computers of staff. We applied the following filtering criteria to select the relevant devices representing the crowd. Other devices are ignored. The signal itself does not tell anything about the device properties. Properties need to be extracted by analyzing the presence over time. A device is categorized as static devices when it is continuously detectable before, and during the exhibition hours. Additionally, we rejected all devices which appeared only during the business hours but their signal was not recognized by a changing set of scanners over the time.

A received signal from out of boundary is defined as a strong signal which was received by scanners at the boundary of the area but did not actually visit the area itself. Obviously the sensors also detect mobile devices neighbored pathways of the area, as it is not possible to setup an electro-magnetic absorbing WiFi shield around the area of interest. The area of interest which we covered with WiFi-scanners accounted for half of the exhibition floor, having one side open to the other half. We grouped scanners into six rows in parallel. Visitors standing not within the

area of interest should have many scans from the boundary scanner rows compared to inner scanners. For each visitor we compute the distribution of scans amongst the six scanner rows and cluster visitors based on a similar scanning pattern. With hierarchical clustering we selected two clusters: devices within the boundary and devices outside the boundary.

During the total experiment duration we scanned 987 681 unique devices, while 317 071 devices were either rejected as being stationary or being out of bounds. This number includes recurring devices over multiple days as the MAC hashing algorithm is changing on each new day.

**Reference Method of Visit Periods** The ‘visit periods’ method serves as a baseline reference for further evaluation comparisons. Similar to entering/exiting counting methods [38, 71, 53] of visitors this method monitors present visitors within the area. This method differs from the other proposed methods as it requires specific knowledge of a) the presence of each specific device and b) is only applicable for the aggregated not local crowd density estimation as the visit period detection within each local area is infeasible with the opportunistic highly varying wireless signal intervals. Both WiFi and Bluetooth scans are calculated separately and are independent features. While a visitor is roaming through the meshed scanner area, any scanner is contributing to pick up his mobile device signal. The scan data from the meshed scanners is aggregated to monitor the presence of a device  $d$  within the area (area covered by the range of the meshed scanners excluding rejected pre-processed devices), by observing the state of appearance  $a$  and disappearance due to time-out  $b$  by any of the aggregated scanners. The time-out of a device is defined by 20 minutes which is the longest silent period discovered by related work [58]. A device  $d$  can be present at multiple visits  $n$ . A visit presence is determined by observing all devices in parallel. The value at time  $t$  is defined as  $\sum\{d \mid a_d n \leq t < b_d n\}$ . The method is computing a time series of active device visit periods which is later used as feature in machine learning. A new computation is performed every 10 seconds, and updates the active device visit periods by adding newly arriving device identifiers and removing previous device identifiers which were observed as gone after the time-out. Within the updated list each device identifier holds a timestamp of the last presence detection. On each update the

last presence detection is compared to the current time stamp and is removed if the time-out of 20 minutes is reached.

**Time Window Based Concept** Towards the aim of local crowd density estimation we introduce the time window based concept which does not rely on historical knowledge of updated visit periods, but only on signals within small windows time windows. We later compare the visit period method to the time window based concept methods. The concept does not rely on tracking the visit period of each device, but relies just on observed devices within the current small time window (without a historic memory). The probability of a device being detected is increasing with an extended observation time window. However, with an increasing time window the crowd density aggregation period is extended which is relevant to changing crowd density situations. The concept methods are computed on a variation of time windows suitable for real-time analysis ranging from 2.5, 5, 10, 20, and 30 minutes. A new computation is performed every 10 seconds allowing real-time estimations, while accessing scan data from the data base of the last 2.5, 5, 10, 20, and 30 minutes.

**Proposed Time Window Based Methods** For the aggregated crowd density estimation we present five methods based on the time window based concept. Each method is computing a measurement value for each 10 seconds of the data set. In the following we present the definition of the methods and highlight their attributes. The proposed crowd density estimation time window based methods are:

- unique WiFi devices per time window (2.5, 5, 10, 20, and 30 minutes)
- unique Bluetooth devices per time window (2.5, 5, 10, 20, and 30 minutes)
- dynamic WiFi devices per time window (2.5, 5, 10, 20, and 30 minutes)
- scans WiFi (2.5, 5, 10, 20, and 30 minutes)
- scans Bluetooth (2.5, 5, 10, 20, and 30 minutes)

The feature *unique WiFi devices per time window (WiFi and Bluetooth)* is calculated for the WiFi and Bluetooth scan data separately. We utilize the fact that most scans contain the unique MAC addresses of devices (uniquely hashed per day). Randomized/dynamic MAC addresses are rejected and not used for this method.

Randomized MAC addresses were marked during the scan phase by identifying a random MAC address pattern not mapping to any vendor identifier in the IEEE MAC vendor lookup table. For each 10 second interval the scans are retrieved from the data set with the selection parameter of the observing time window of  $t_{-2.5}...t_0$ ,  $t_{-5}...t_0$ ,  $t_{-10}...t_0$ ,  $t_{-20}...t_0$ , and  $t_{-30}...t_0$ . For the aggregated crowd density estimation no further selection based on the scanner identifier is made. A set is created with all hashed MAC addresses occurring in the observed time window. Finally, the size of the set represents the number of unique MAC addresses and denotes one value within the time series of one observing time window. The same procedure is repeated for each time window.

The feature *dynamic devices per time window (WiFi)* builds on the existence of randomized MAC addresses [8] and their observations [110]. This method is considering the potential change towards full WiFi probe MAC randomization and the effect on crowd density estimation. This method is not applied to Bluetooth signals as MAC randomization is not applied to Bluetooth addresses. Device vendors are shifting towards sending dynamic WiFi probe MAC addresses for privacy reasons. This is currently the case for WiFi MAC addresses. The time interval of dynamic MAC address change is not specified by the device vendors and is expected to vary. These dynamic MAC addresses are identifiable as not complying to IEEE vendor data base entries. In related work dynamic MAC addresses are often not identified and lead to confusion. We recognize dynamic MAC addresses and calculate the set size of distinct dynamic MAC addresses similar to the previous method *unique WiFi devices per time window (WiFi and Bluetooth)* but only selecting randomized MAC address scan entries.

The feature *number of total scans (WiFi and Bluetooth)* builds on the assumption of an increasing number of scans with an increasing number of present devices. For each 10 second interval the scans are retrieved from the data set with the selection parameter of the observing time window of  $t_{-2.5}...t_0$ ,  $t_{-5}...t_0$ ,  $t_{-10}...t_0$ ,  $t_{-20}...t_0$ , and  $t_{-30}...t_0$ . For the aggregated crowd density estimation no further selection based on the scanner identifier is made. Resulting scan entries can involve an arbitrary number of scans per device along all scanners. Repeating scans by the same scanner are included, and repeating scans of the same device by different scanners are also included. This method is fully independent of unique and dynamic/randomized

MAC addresses as only the absolute size of retrieved scans is considered by this method. Finally, the size of the returned scans fulfilling the observation time window criterion denote one value within the time series of one observing time window. The same procedure is repeated for each time window.

**Distinction to Other Time Series Based Analysis Methods** Several specific time series processing methods exist which benefit from ordered data sequences. Popular methods are Bernoulli process, Markov chain, random walk, maximal entropy walk, or the hidden Markov model. One of the most popular method is the hidden Markov model where the current state is estimated by probability transition i.e. trained with the Baum–Welch algorithm to find the unknown parameters of a hidden Markov model. This method is often used in the context of activity recognition. The methods were considered but due to the continuity of the crowd density estimation during the experiments the methods were not further regarded because stronger variations in the crowd density would have been necessary to evaluate the effect of such time series based estimation methods.

**Machine Learning Algorithm Overview** As stated before, the actual fraction of visitors carrying a WiFi enabled is not known a-priori. The knowledge needs to be extracted from the data set, with the extracted features and the ground truth values. We apply multiple popular machine learning algorithms applicable to the regression problem, and later evaluate the machine learning algorithms towards their estimation error. In general, supervised machine learning is based on extracted feature vectors and ground truth target variables, while training the regressor with a small part of the data set and validating the regressor with the remaining part of the data set. We applied the following machine learning algorithms which are briefly described afterwards:

- linear regression with ordinary least squares
- kernel ridge regression
- support vector regression
- gaussian process regression
- regression trees
- neural network based regression

Linear regression with ordinary least squares is an attractive model because the representation is very simple. Linear regression fits a linear model with the coefficients  $c = (c_1, \dots, c_n)$  to minimize the residual sum of squares between the observed responses in the dataset, and the responses predicted by the linear approximation. Mathematically, it solves a problem of the form:  $\min_c \|Xc - y\|^2$ . Kernel ridge regression combines ridge regression with the kernel trick. It learns a linear function by the respective kernel and the data. Ridge regression addresses some of the problems of ordinary least squares by creating a penalty on the size of coefficients. The ridge coefficients minimize a weighted residual sum of squares,  $\min_c \|Xc - y\|^2 + \alpha \|c\|^2$ .  $\alpha \geq 0$  is a parameter that controls the amount of shrinkage: the larger the value of  $\alpha$ , the greater the amount of shrinkage and thus the coefficients become more robust to collinearity. The support vector regression with the linear kernel is an extension to the support vector machine classification algorithm. A support vector machine constructs a hyperplane to discriminate between instances. Gaussian process regressors are another machine learning algorithm to solve a regression problem with a probabilistic approach in kernel machines. The kernel '1.0 \* RBF(1.0)' is used. The kernels hyper-parameters are optimized during fitting of the Gaussian process regressor by maximizing the log-marginal-likelihood. Regression trees are dependent from classification decision trees. As in the classification setting, the training phase takes as input the feature vector  $X$  and target variable  $y$ , only that in this case  $y$  is expected to have floating point values instead of integer values. A popular neural network algorithm is the multi-layer perceptron that trains using back propagation with no activation function in the output layer. It uses the square error as the loss function, and the output is a set of continuous values.

**Machine Learning Training/Validation Overview** We evaluate all features individually on the machine learning algorithms. To evaluate the machine learning algorithms we selected the common 10-fold cross-validated for dividing the data set in separate training and validation sets, while each element is once used for validation.

### 2.8.3 Aggregated Crowd Density Estimation Results

In this sub-section the results of the machine learning validation based on the previously defined methods are presented, including the description of the results of the visit period method for all machine learning algorithms, the detailed linear regression result description of individual days. We divide the presentation of the crowd density estimation results into the WiFi based results and the Bluetooth based results. The following described results of the methods are shown in Table 2.3. Firstly, we discuss the results of the reference method ‘visit periods’ (WiFi) and present them in the order of the ascending error. The error has to be seen in the context of the average crowd density of 0.22 during the experiment. The kernel ridge regression has a mean absolute error (MAE) of 0.1070 [unit of people per  $m^2$ ]. The linear regression has a MAE of 0.0112. The decision tree regression has a MAE of 0.0155. The support vector regression has a MAE of 0.0190. The gaussian process regression has a MAE of 0.0287. The neural network based regression has a MAE of 0.0445. The median error (MMAE) is lower for all algorithms which is caused by the fact that higher crowd densities are predominating and having a better statistical expressiveness.

Secondly, we present the results of the proposed time window based methods (WiFi) in the order of the ascending error for the linear regression algorithm. The method ‘unique WiFi devices per 2.5 minute time window’ (‘unique 2.5 min WiFi’) has the lowest estimation error of 0.0115 beyond the reference method. Followed by the same method with increasing time windows with increasing MAE of up to 0.0153. The method ‘dynamic’ follows with MAE of 0.0196–0.0199 for all time windows. The method ‘scan’ follows in reverse order of the time window size ranging from MAE 0.0393 for ‘scan 30 min’ to MAE 0.0497 for ‘scan 2.5 min’. Other machine learning algorithms produce different results. The kernel ridge regression algorithms results in larger MAE, while the distribution of the errors between time windows is different than with linear regression. The support vector regression algorithms results in slightly smaller MAE than linear regression for many of the methods. The gaussian process regression algorithm results in overall higher MAE than linear regression. The regression decision trees algorithm in slightly higher MAE with a minor number of methods having a lower MAE than linear regression. The



neural network based regression algorithm results in overall higher MAE than linear regression, which is caused by the disadvantage of neural networks with an insufficient number of training elements.

Thirdly, we discuss the results of the reference method ‘visit periods’ (Bluetooth) and present them in the order of the ascending error. The support vector regression has a mean absolute error (MAE) of 0.0463 [unit of people per  $m^2$ ]. The gaussian process regression has a MAE of 0.0468. The neural network based regression has a MAE of 0.0497. The linear regression has a MAE of 0.0550. The decision tree regression has a MAE of 0.0554. The kernel ridge regression has a MAE of 0.1597. The median error (MMAE) is lower for all algorithms which is caused by the fact that higher crowd densities are predominating and having a better statistical expressiveness.

Fourthly, we present the results of the proposed time window based methods (Bluetooth) in the order of the ascending error for the linear regression algorithm. The method ‘unique Bluetooth devices per 30 minute time window’ (‘unique 30 min Bluetooth’) has the lowest estimation error of 0.0456, even lower than the reference method. Followed by the same method with decreasing time windows with increasing MAE of up to 0.0488. The method ‘scan’ follows with a MAE of 0.0456 for ‘scan 2.5 min’, followed by the remaining time windows in reversed order. With all machine learning algorithms the Bluetooth methods result in higher MAE than WiFi based methods.

**Table 2.3.** Crowd density estimation errors [unit of people per  $m^2$ ] with machine learning evaluation of the reference method ‘visit periods’ and the proposed time window based methods. The machine learning algorithms include the linear regression with ordinary least squares (Lin. Reg.), the kernel ridge regression (Kernel Ridge), the support vector regression (SVR Reg.), the gaussian process regression (GP Reg.), the regression decision trees (DT Reg.), and the neural network based regression (Neural Net). The presented estimation error are the mean absolute error (E1) and the median absolute error (E2). Both wireless scan methods of Bluetooth and WiFi are presented

Feature	Wireless	Lin. Reg.		Kernel Ridge		SVR Reg.		GP Reg.		DT Reg.		Neural Net	
		E1	E2	E1	E2	E1	E2	E1	E2	E1	E2	E1	E2
Visit periods	WiFi	0.0112	0.0094	0.1070	0.1050	0.0190	0.0142	0.0287	0.0203	0.0155	0.0126	0.0445	0.0332
Unique 2.5 min	WiFi	0.0115	0.0094	0.0962	0.0950	0.0144	0.0115	0.0247	0.0179	0.0151	0.0123	0.0462	0.0363
Unique 5 min	WiFi	0.0124	0.0100	0.0722	0.0712	0.0121	0.0101	0.0191	0.0146	0.0150	0.0120	0.0429	0.0360
Unique 10 min	WiFi	0.0134	0.0110	0.0471	0.0473	0.0113	0.0095	0.0144	0.0115	0.0141	0.0112	0.0373	0.0331
Unique 20 min	WiFi	0.0153	0.0128	0.0278	0.0274	0.0120	0.0100	0.0120	0.0097	0.0136	0.0104	0.0352	0.0280
Unique 30 min	WiFi	0.0153	0.0127	0.0219	0.0217	0.0120	0.0097	0.0116	0.0086	0.0131	0.0097	0.0399	0.0362
Dynamic 2.5 min	WiFi	0.0196	0.0172	0.1493	0.1458	0.0398	0.0279	0.0416	0.0290	0.0221	0.0180	0.0482	0.0353
Dynamic 5 min	WiFi	0.0199	0.0174	0.1317	0.1294	0.0297	0.0211	0.0347	0.0243	0.0217	0.0175	0.0467	0.0356
Dynamic 10 min	WiFi	0.0199	0.0170	0.0949	0.0925	0.0185	0.0154	0.0243	0.0180	0.0209	0.0172	0.0444	0.0350
Dynamic 20 min	WiFi	0.0198	0.0170	0.0499	0.0488	0.0158	0.0135	0.0177	0.0151	0.0200	0.0164	0.0404	0.0352
Dynamic 30 min	WiFi	0.0195	0.0170	0.0311	0.0277	0.0153	0.0130	0.0166	0.0144	0.0194	0.0158	0.0333	0.0314
Scans 2.5 min	WiFi	0.0497	0.0472	0.0498	0.0472	0.0283	0.0221	0.0360	0.0198	0.0286	0.0215	0.1193	0.0928
Scans 5 min	WiFi	0.0456	0.0443	0.0456	0.0443	0.0319	0.0248	0.0461	0.0217	0.0244	0.0185	0.1819	0.1549
Scans 10 min	WiFi	0.0427	0.0421	0.0427	0.0421	0.0309	0.0253	0.0677	0.0328	0.0203	0.0157	0.4104	0.3720
Scans 20 min	WiFi	0.0404	0.0399	0.0404	0.0399	0.0286	0.0228	0.0935	0.0656	0.0178	0.0139	0.6507	0.6623
Scans 30 min	WiFi	0.0393	0.0385	0.0393	0.0385	0.0349	0.0224	0.1092	0.0920	0.0163	0.0126	1.1282	0.9666
Visit periods	Bluetooth	0.0550	0.0396	0.1597	0.1558	0.0463	0.0326	0.0468	0.0327	0.0554	0.0363	0.0497	0.0365
Unique 2.5 min	Bluetooth	0.0488	0.0338	0.1546	0.1505	0.0463	0.0326	0.0468	0.0327	0.0586	0.0399	0.0497	0.0362
Unique 5 min	Bluetooth	0.0486	0.0335	0.1516	0.1473	0.0462	0.0325	0.0467	0.0326	0.0574	0.0386	0.0490	0.0356
Unique 10 min	Bluetooth	0.0479	0.0330	0.1463	0.1421	0.0458	0.0322	0.0466	0.0325	0.0552	0.0372	0.0491	0.0352
Unique 20 min	Bluetooth	0.0465	0.0321	0.1367	0.1328	0.0452	0.0318	0.0461	0.0322	0.0521	0.0335	0.0496	0.0376
Unique 30 min	Bluetooth	0.0456	0.0319	0.1279	0.1240	0.0446	0.0313	0.0456	0.0318	0.0507	0.0319	0.0504	0.0376
Scans 2.5 min	Bluetooth	0.0456	0.0319	0.1279	0.1240	0.0446	0.0313	0.0456	0.0318	0.0507	0.0319	0.0492	0.0370
Scans 5 min	Bluetooth	0.0530	0.0360	0.0542	0.0337	0.0431	0.0289	0.0437	0.0301	0.0579	0.0363	0.0531	0.0430
Scans 10 min	Bluetooth	0.0522	0.0352	0.0504	0.0314	0.0432	0.0290	0.0443	0.0319	0.0582	0.0367	0.0606	0.0500
Scans 20 min	Bluetooth	0.0517	0.0349	0.0510	0.0337	0.0430	0.0284	0.0456	0.0348	0.0577	0.0360	0.0643	0.0553
Scans 30 min	Bluetooth	0.0509	0.0342	0.0507	0.0339	0.0432	0.0287	0.0467	0.0356	0.0555	0.0333	0.0497	0.0365

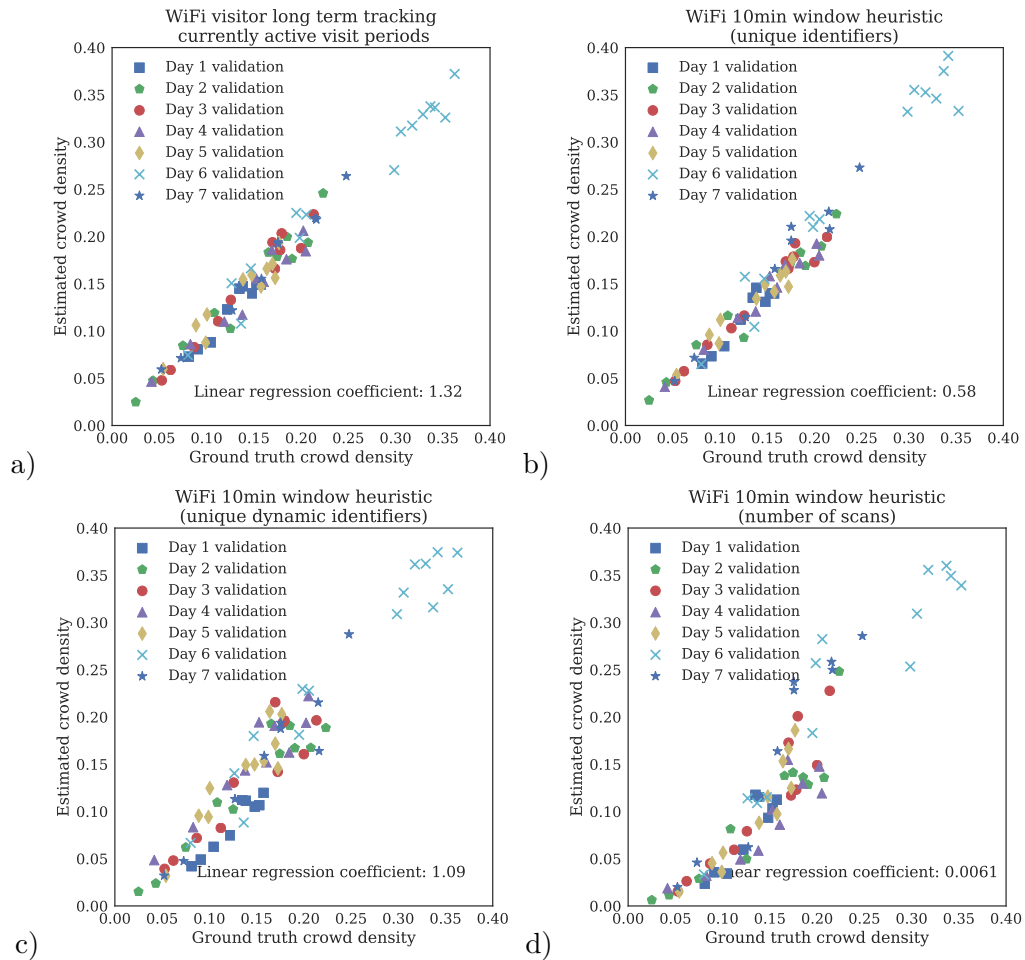
**Linear Regression Analysis** The linear regression algorithm results in the second best MAE rates over all machine learning algorithms. The advantage of linear regression -when a linear connection is existing- is the easily comprehensible conversion from input data to output data. In this paragraph we discuss the results of the linear regression algorithms for multiple experiment days. Linear regression with the method ‘visit periods’ (WiFi) has a coefficient of 1.32 based on seven experiment days including different private or business backgrounds of visitors. The coefficient of 1.32 equals to 76 % of visitors being equipped with WiFi enabled devices. The method ‘unique 10 min (WiFi)’ has a coefficient of 0.58, which means every device is discovered 1.7 times within 10 minutes independent of the actual visit duration. The feature ‘dynamic 10 min (WiFi)’ has a coefficient of 1.09. The feature ‘scan 10 min (WiFi)’ has a coefficient of 0.0061.

The resulting linear regression Figure 2.7 shows the resulting connection between the ground truth crowd density and the estimated crowd density for the individual experiment days. The reference method ‘visit periods (WiFi)’ is shown in Figure 2.7a), the ‘unique 10 min (WiFi)’ method in Figure 2.7b), the ‘dynamic 10 min (WiFi)’ method in Figure 2.7c), and the ‘scan 10 min (WiFi)’ method in Figure 2.7d). We see that the variation of the estimation increases proportionally with increasing crowd density for all methods. The method ‘unique 10 min (WiFi)’ has very similar ‘visit periods (WiFi)’ (see Figure 2.7a and b). The variation in the estimation is significantly larger with method ‘dynamic 10 min (WiFi)’ (see Figure 2.7c) across all experiment days, while a linear relationship is still observable. The method ‘scans 10 min (WiFi)’ performs significantly worse with linear regression (see Table 2.3) as linear relationship is not observable (see Figure 2.7d), which is also reinforced by the fact that other methods perform better with the method ‘scan’, such as the decision tree regression algorithm with a MAE of 0.0203 compared to the linear regression MAE of 0.0427 (see Table 2.3).

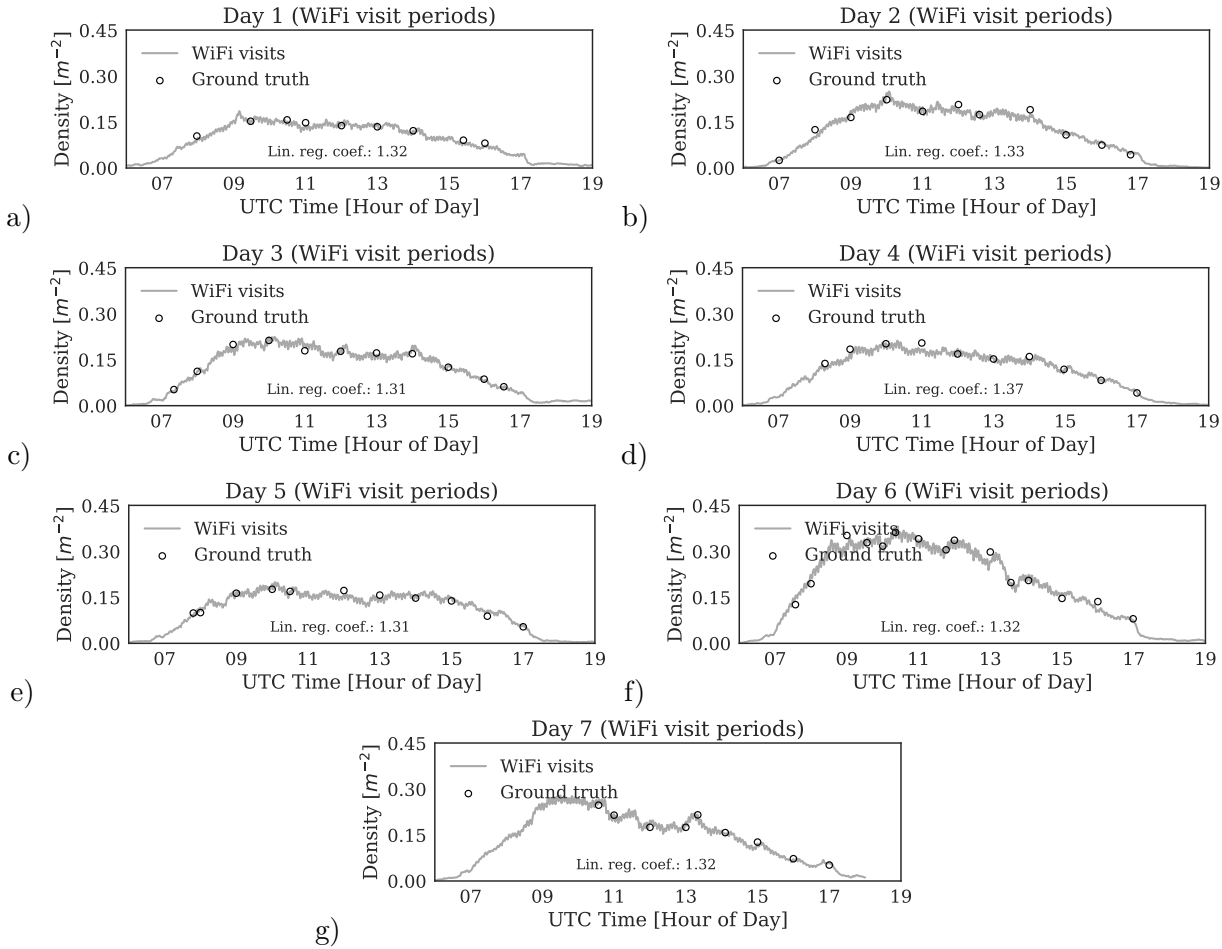
Figure 2.8 is showing the resulting estimated crowd density time series by the method ‘unique 10 min (WiFi)’ along with the ground truth for each experiment day. We observe a reliable estimation even when the crowd density fluctuates over a short period of time. Such fluctuations are especially observable on experiment day 6 and experiment day 7 (see Figure 2.8f and g).

Differently to WiFi methods, the number of detected Bluetooth fluctuates heavily

over time. Linear regression with the feature ‘visit periods (Bluetooth)’ has a coefficient of 6.67 based on seven experiment days during different crowd densities including different private or business backgrounds of visitors. The method ‘unique 10 min (Bluetooth)’ has a coefficient of 3.09. The method ‘scan 10 min (Bluetooth)’ has a coefficient of 0.11.



**Figure 2.7.** Linear regression algorithm results showing the relation between the ground truth crowd density and the estimated crowd density for the individual experiment days. The regression results are shown for a) the reference method ‘visit periods (WiFi)’, b) the ‘unique 10 min (WiFi)’ method, c) the ‘dynamic 10 min (WiFi)’ method, and d) the ‘scan 10 min (WiFi)’ method.



**Figure 2.8.** Time series of the estimated crowd density during seven days of the experiment. Includes ground truth annotation values. The estimation results are based on the selected linear regression algorithm with the method ‘unique 10 min (WiFi)’.

## 2.8.4 Methods and Results of Local Crowd Density Estimation

In the previous section we demonstrated the feasibility of the aggregated crowd density estimation with evaluations based on the area-wide ground truth. In this section we describe the transfer of the methods to the local crowd density estimation, including the definition of the local crowd density scale, the Voronoi

cell binning, the method, the calibration approach of the local crowd density based on the global crowd density, the calibrations coefficients, and finally describe the qualitative results for local crowd density estimation visualization together with visual ground truth.

As stated before, the ‘local’ crowd density scale is referring to the area of one scanning zone and still represents the averaged number of people per square meter but within a much smaller scale of in average  $180m^2$ , which corresponds to a circular area with a radius of just 7.5 m or to a quadratic area with a width of just 13.5 m. This scale is useful when analyzing the distribution of the crowd density within a two-dimensional visualization, which is essential when analyzing which regions are more densely packed than others. The scanners are carefully set up to cover the whole area without leaving ‘dead-zones’ in scanning of the devices. Knowledge has been collected from previous experiments and brought into the setup of the scanners. Conventional opportunistic smartphone signals (other wireless signals are stronger and need to be rejected by pre-processing) are detected within the proximity of one scanner with more infrequent detections by neighbored scanners. We define the region of each scanner as its Voronoi cell. A Voronoi cell has the attribute that each point within the cell is closer to its scanner than to any other scanner. Essentially, the scanner setup defines the scale of the crowd density measurement. However, decreasing the scanner displacement and therefore decreasing the scanner cell size is not feasible because of the impact of signal attenuations induced by the crowd and not only due to the distance.

We transfer the previous aggregated method ‘unique scan’ to the local method ‘local unique scan’. We rely on the WiFi methods which have been previously proven more reliable in this scenario. Randomized/dynamic MAC addresses are rejected and not used for this method. For each 10 second interval the scans are retrieved from the data set with the selection parameter of the observing time window of  $t_{-2.5}...t_0$ ,  $t_{-5}...t_0$ ,  $t_{-10}...t_0$ ,  $t_{-20}...t_0$ , and  $t_{-30}...t_0$ . For the local crowd density estimation we select the scans by each scanner identifier from the base separately. A set is created with all hashed MAC addresses occurring in the observed time window for. Finally, the size of the set represents the number of unique MAC addresses and denotes one value within the time series of one observing time window. The same procedure is repeated for each time window and for each scanner. When

data is selected by one scanner this happens in isolation to scans of the same device identifiers by other scanners. This means that a roaming visitor device can be detected by multiple scanners within one time window and support the crowd density estimation within multiple cells. This is a correct assumption as roaming visitors also contribute to the local crowd density for a short period of time. The probability of detecting a roaming visitor within a cell is increasing with the duration spent within a cell. This assumption is supported by the large data set and the large number of detected devices. We choose the linear regression algorithm for the local crowd density estimation. To determine the coefficient for local crowd density estimation given the aggregated crowd density ground truth we rely on the previous results and derive the coefficient based on the averaged local method values over all cells which is compared to the aggregated crowd density.

The resulting linear regression coefficients for the method ‘local unique scan’ are shown in Table 2.4 for multiple time windows. Based on the extracted coefficients the method ‘local unique scan’ is calibrated and used for further visualization of the local crowd density.

**Table 2.4.** Coefficient values for different time windows based on supervised linear regression machine learning validation and the ‘unique’ features for *local* Voronoi cell based crowd density estimation.

Time Window	Coefficient ‘unique’
2.5 min	1.4788
5.0 min	1.0819
10.0 min	0.0657
20.0 min	0.4907
30.0 min	0.3749

In Figure 2.9 and Figure 2.10 we present snapshots of the local crowd density scale estimation as heat-map visualizations. Blue cells denote low crowd densities, white cells denote average crowd densities, and red cells denote high crowd densities. The visualization includes the scanner locations, the objects of interest, the Voronoi cell borders, and obstructions such as walls. The series of heat-maps represent

the filling phase and stagnant phase during the experiment day 2 which is also presented as the time series of the aggregated crowd density in Figure 2.8b. The filling phase is consisting of the first four hours of the business day 2. The crowd density stagnates at around 10 am (UTC). Then small crowd density displacements are observable over time. Spatially, the highest crowd density is observed within the center of the area where an exhibit (world premiere show car) was located which attracted many visitors at a time and led to a intermittent local crowd density of up to 0.45 people per square meter, which is twice as much as the aggregated crowd density of 0.22 during this time.

Next to the heat-map we present the ground truth images at the according point in time. Especially at the center of the ground truth images we can observe that the area around the world premiere show car attracts many visitors which aligns with the heat-map visualizations. The left, top, and right border of the area are hard to perceive in the small images ground truth images. Exact analysis together with knowledge gathered during the event affirms the given crowd density distribution shown in the heat-maps.



### **2.8.5 Methods and Results of Local Crowd Movement Estimation**

In the previous sections we demonstrated the feasibility of the aggregated and local crowd density estimation scale. In this section we describe the crowd movement approach, the Voronoi cell binning, and finally present the qualitative results for local crowd movement estimation by visualization with visual ground truth.

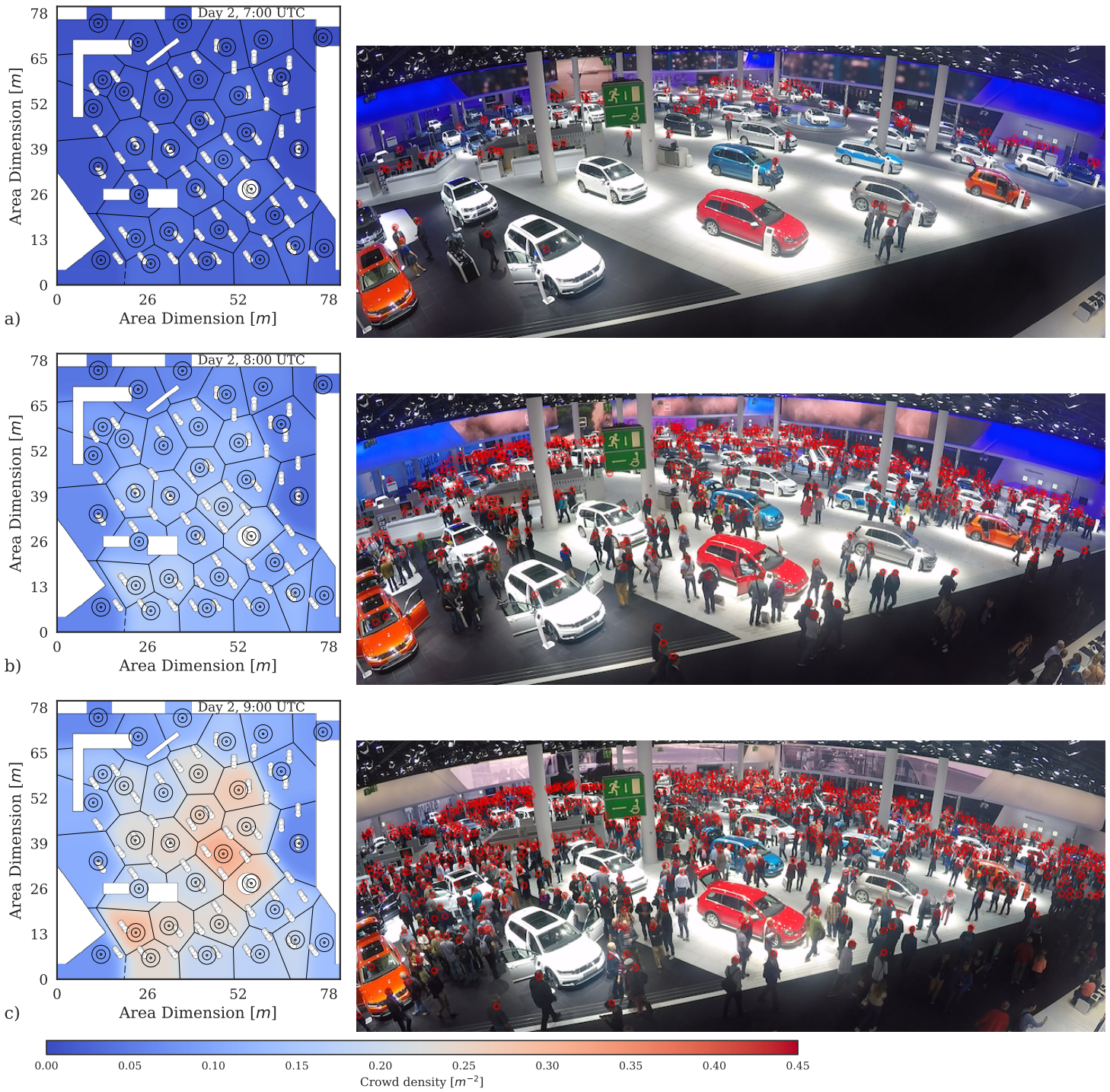
In general, the ‘crowd movement’ refers to the average velocity of people within a crowd. In related work the crowd movement has been determined as an import crowd condition measurements as critical crowd movement levels have been demonstrated where the crowd movement begins to decrease as the crowd density rises to a critical level [33]. We present the local crowd movement as a value being a relative indicator within the experiment. This is based on the fact of insufficient ground truth camera angles to reliably determine the visitor path in all areas due to intermittent obstruction of visitors in the visual ground truth video stream. The evaluation would not resilient regarding the ground truth. However, we present the crowd movement visualizations and qualitative results compared to best-effort crowd movement ground truth extractions.




The fundamentals of the crowd movement method is similar to the local crowd density estimation in subsection 2.8.4, including the Voronoi cell binning and the time window based approach. We propose a method ‘unique once assign’ for the crowd movement estimation by assigning a device within a time window once to a single cell. Randomized/dynamic MAC addresses are rejected and not used for this method. For each 10 second interval the scans are retrieved from the data set with the selection parameter of the observing time window. For the local crowd movement estimation we select the scans by each scanner identifier from the data base separately. The method then calculates the number of scans of each scanner. Each device is then assigned to the cell with the highest count of scans. This process is repeated for all devices within this time window. Finally, the assigned devices per cell are accumulated for each cell. Differently to the previous method ‘local unique scan’ each device contributes only exactly once to a scanner cell and not to multiple cells while the user is roaming through the area. When a visitor is continuously roaming this could be any cell within the path. When a visitor is standing -and

scans are accumulated- during the time window the probability is increased of the local scanner cell to be selected by the algorithm. The methods builds on the assumption of uniformly distributed WiFi probe requests and Bluetooth inquiry responses during the visit. Having a big data set containing thousands of device scans in parallel, the method also builds on the assumption that different visitors will be scanned by different scanners during their visit event when having the same path.

In Figure 2.11 we present snap-shots of the local crowd movement estimation as heat-map visualizations. Blue cells denote high crowd flow, white cells denote average crowd flow, and red cells denote low crowd flow. The visualization includes the scanner locations, the objects of interest, the Voronoi cell borders, and obstructions such as walls. The series of heat-maps represent the same points in time as in Figure 2.9 and Figure 2.10. It is observable that the crowd movement is changing over time. Spatially, the lowest crowd flow is observed within the center of the area where an exhibit (world premiere show car) was located which attracted many visitors at a time. Along the heat-maps we present the ground truth images at the according point in time. At the center of the ground truth images we can observe that the area around the world premiere show car attracts standing visitors -contributing to low crowd movement- which aligns with the heat-map visualizations.

2 Crowd Condition Estimation with Stationary Scanners Opportunistically Scanning Crowd Devices

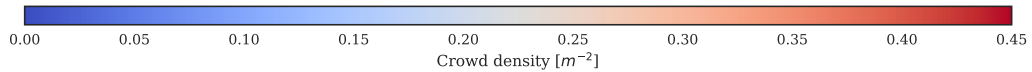
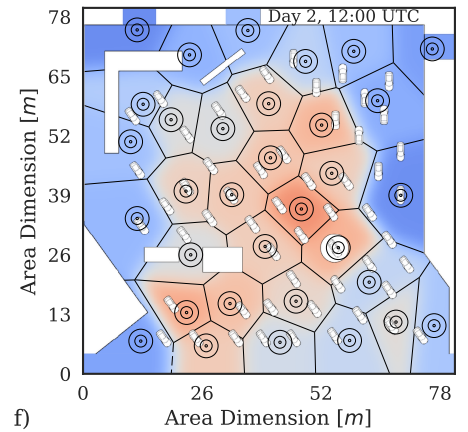
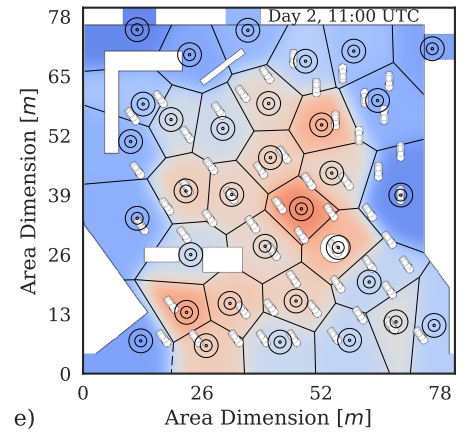
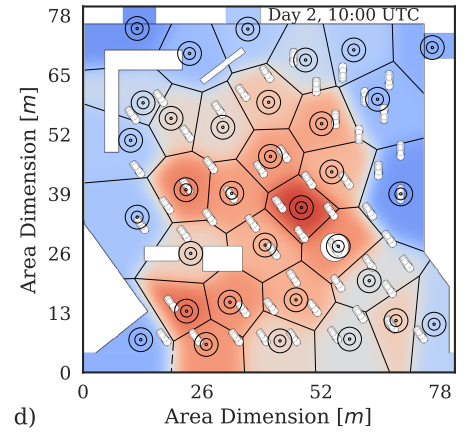





 Object of visitor interest  
  Scanner unit (placement at ceiling)  
  Walls/obstructions

**Figure 2.9.** Heat-map visualization of the crowd density based on the local crowd density estimation method. Snap-shots (density evolves over 10 minutes) of the crowd density is shown at 07:00am UTC, 08:00am, and 09:00am during the filing phase of the second experiment day. See Figure 2.8b for the aggregated crowd density during the same day. Red circles in ground truth image show manual human annotations for machine learning training.



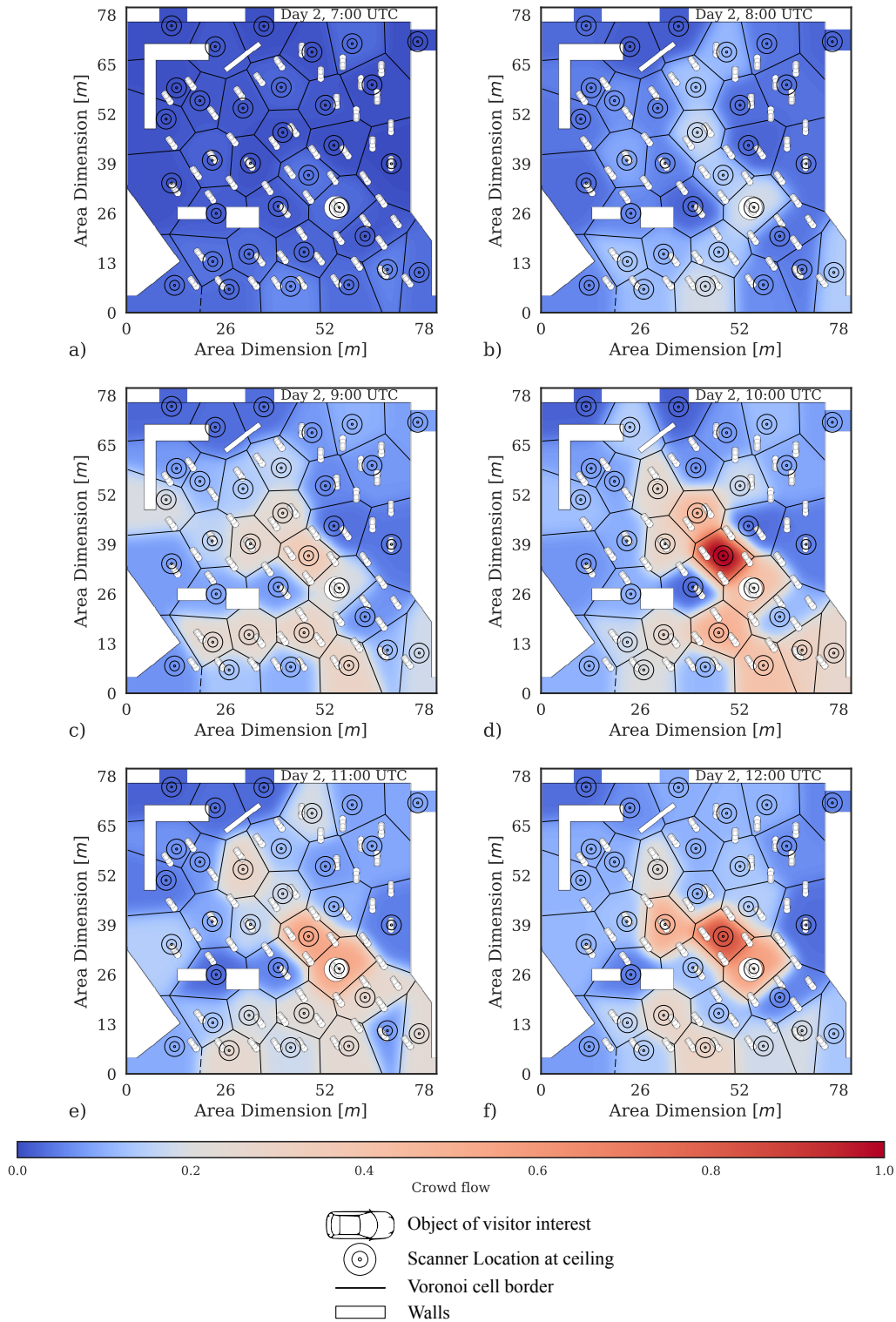
## 2.8 Methods and Results



 Object of visitor interest    
  Scanner unit (placement at ceiling)    
  Walls/obstructions

**Figure 2.10.** Heat-map visualization of the crowd density based on the local crowd density estimation method. Snap-shots (density evolves over 10 minutes) of the crowd density is shown at 10:00am, 11:00am, and 12:00am during the stagnant phase during the second experiment day. Red circles in ground truth image show manual human annotations for machine learning training. See Figure 2.8b for the aggregated crowd density during the same day.

2 Crowd Condition Estimation with Stationary Scanners Opportunistically Scanning Crowd Devices



58

**Figure 2.11.** Heat-map visualization of the crowd movement estimation within cells of the meshed scanners. Visualized for multiple points in time at 07:00am UTC, 08:00am, 09:00am, 10:00am, 11:00am and 12:00am during the filing phase. The crowd movement level varies between the daily range values 0.0 and 1.0 along the scanner cells.

### 2.8.6 Methods and Results of Contextual Crowd Condition Anomaly Detection

In the previous section we presented the crowd density estimation methods and crowd movement methods. In this section we present contextual crowd condition anomaly detection, including methods and qualitative evaluations based on two significant events (visit of the German chancellor Dr. Merkel and the press conference) during the experiment.

Crowd condition estimation and crowd movement estimation is suitable for continuous real-time and online monitoring. However, towards full insight into the crowd condition besides absolute crowd density and crowd movement the temporal course is important. For example, a high crowd density is necessary for a critical situation but not compulsory. We define the contextual monitoring of the crowd condition as monitoring the change of measurements over time. A potential finding in the course of the crowd measurements are called anomalies. Anomalies are divergences from the normal/usual course in time. An anomaly can be a rapid increase or decrease in the crowd density, changes in the crowd flow or other changes from the crowd behavior. We present methods for crowd condition anomaly detection. We do not provide a summary on general anomaly detection categories such as point anomalies, contextual anomalies or collective anomalies used for system intrusion detection, fraud detection or fault/damage detection, i.e. surveyed by Chandola et al. [19]. We rather present the central ideas of time sequence based contextual crowd condition anomaly detection. Related work on crowd condition anomaly is non-existent for wireless signal sensing and is limited to computer vision based crowd condition anomaly detection as in [65].

The methods rely on the time window based concept introduced earlier. The proposed anomaly detection algorithm relies on detecting rapid variations of a short time window compared to a longer overlapping time window. We define a short time window as a duration of 10 minutes motivated by the observation of detecting over 80% of wireless enabled devices within a 10 minute time window [81]. We define the overlapping time window as a duration of 60 minutes. The 60 minute spanning time window is motivated by the background knowledge of anomalies. In our experimental case the anomalies are short term anomalies with a duration

well below 60 minutes. Anomalies lasting longer need a longer spanning time window. However, the thorough analysis of time window sizes is out of the scope of this thesis due to the lack of the existence of anomalies of different lengths during the experiment. The algorithms is defined as follows: 1. the rolling mean (time window of 60 minutes) is applied to the crowd condition measurements, 2. subtracting the small 10 minute window crowd condition measurements from the previous overlapping 60 minute rolling mean crowd condition measurements. This calculation is performed repeatedly for one minute sliding window steps.

We apply the proposed anomaly algorithm to multiple crowd condition measurements - in addition to the crowd density measurement and crowd flow measurement. The anomaly algorithm and the measurements are briefly described in the following list:

- The anomaly measurement ‘divergence of unique devices’ at time  $t$  is defined as the disparity between measurement ‘aggregated unique 10 min’ ( $t_{-5}$ )... $t_{+5}$ ) to the measurement ‘aggregated unique 60 min’ ( $t_{-30}$ )... $t_{+30}$ ).
- The anomaly measurement ‘divergence of unique arrivals’ is defined as the disparity between the number of arrivals (first occurrences) of unique device identifiers to the number of arrivals within the last 60 minutes.
- The anomaly measurement ‘divergence of unique departures’ is defined as the disparity between the number of departures (minimum of 10 minute time-out since last observation) of unique device identifiers and the number of departures within the last 60 minutes.
- The anomaly measurement ‘standard deviation of unique detections’ is defined as the standard deviation of the measurement ‘aggregated unique 60 min’ at time  $t$ .
- The anomaly measurement ‘standard deviation of unique arrivals’ is defined as the standard deviation of the number of arrivals within the last 60 minutes at time  $t$ .
- The anomaly measurement ‘standard deviation of unique departures’ is defined as the standard deviation of the number of departures within the last 60 minutes at time  $t$ .

The results of the applied anomaly algorithm is presented in Figure 2.12 and

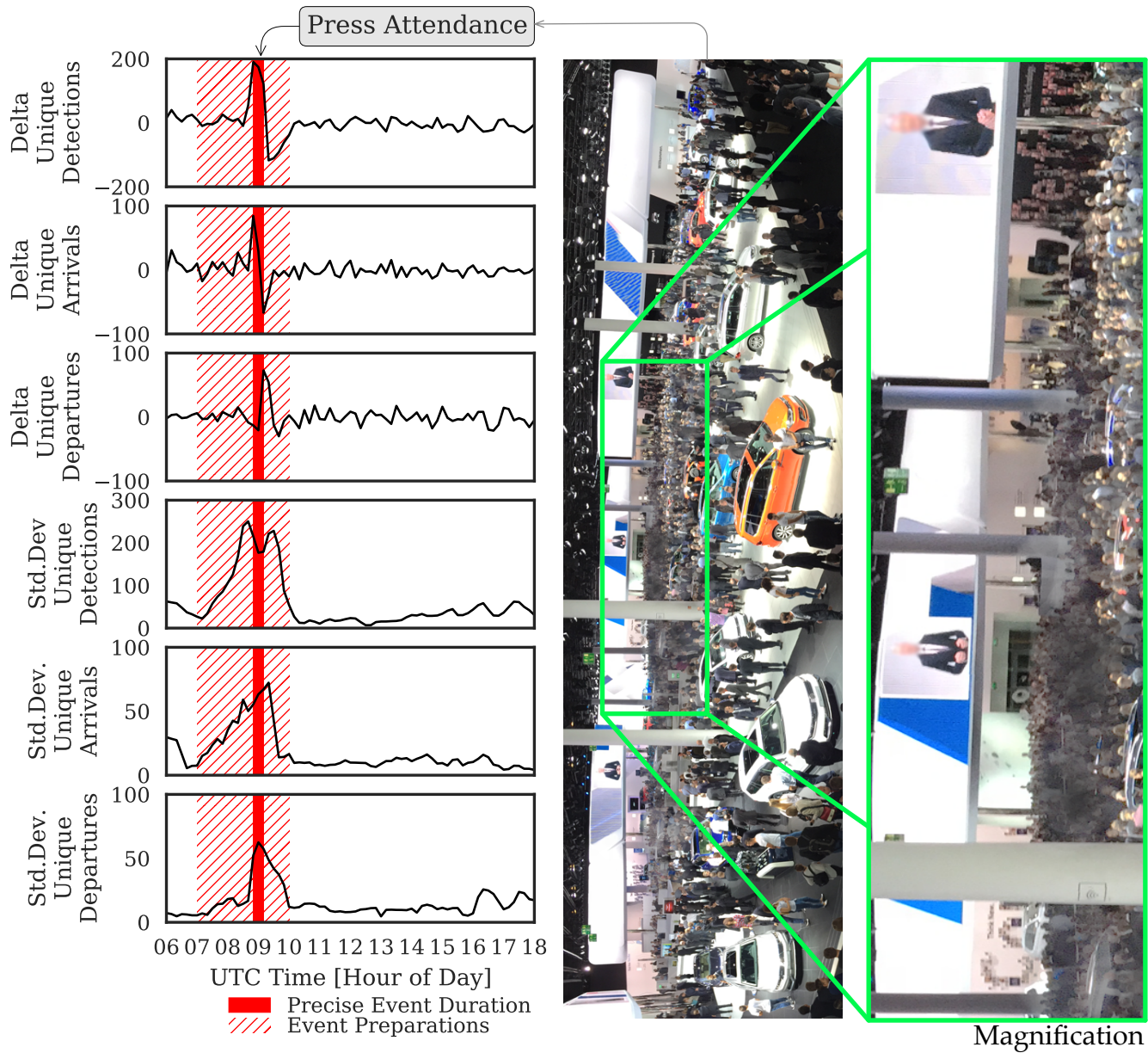


Figure 2.13. Two significant crowd condition anomalies happened during the experiment days: one being a 15 minutes press conference and one being a 20 minutes visit by the German chancellor Dr. Merkel.

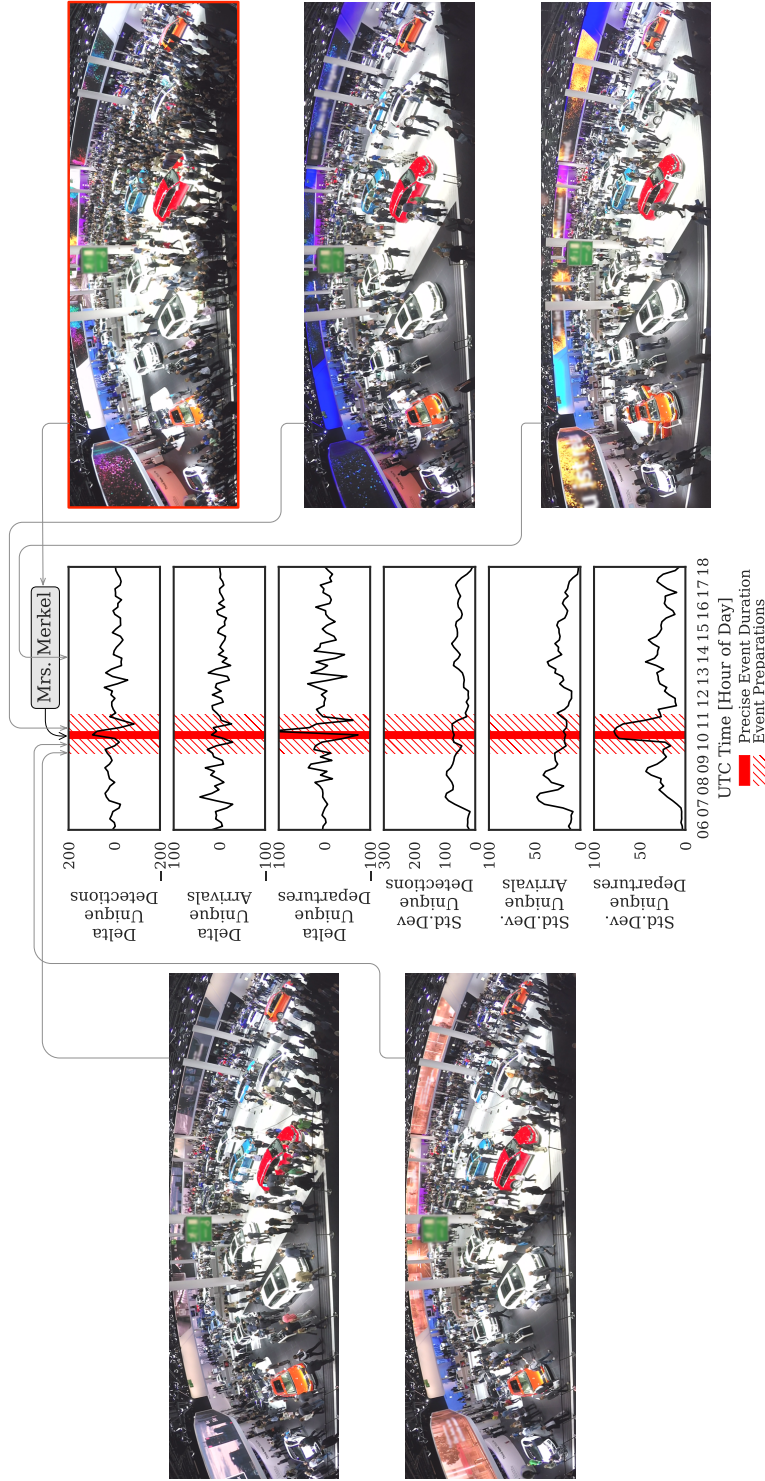
In Figure 2.12 the press-conference is highlighted (from 9:00am to 9:15am UTC) with a prior event preparation phase (i.e. area restructuring and spectators arrival) and relieve phase (i.e. area restructuring, chancellor/staff departure, and spectators departure). During the press-conference we observe peaks in all anomaly algorithm measurements immediately before, during, or immediately after the event. As noted before, automatic detection of such ‘peak’ events is straightforward with common peak detection algorithms, however, a systematic evaluation is infeasible due to the existence of just two proven crowd condition anomalies. The reliability of a valid anomaly can be considered high because all anomaly algorithm measurements show peaks.

In Figure 2.13 the visit of chancellor Dr. Merkel is highlighted (from 10:30am to 10:50am UTC). During the visit of German chancellor Dr. Merkel we observe peaks in anomaly algorithm measurements unique detections, unique departures and standard deviation of unique departures. No peaks are monitored for the anomaly algorithm measurements of delta unique arrivals, standard deviation unique detections and standard deviation unique arrivals.





**Figure 2.12.** Highlighted contextual anomaly detection algorithm results showing the event of the press conference. A contextual anomaly is defined as a temporal divergence from normal/usual course of crowd condition measurements. Showing a high crowd density during the event, fast arrivals before the event, fast departures after the event, and further standard deviation indicators.



**Figure 2.13.** Highlighted contextual anomaly detection algorithm results showing the event of the visit of chancellor Dr. Merkel. A contextual anomaly is defined as a temporal divergence from the normal/usual course of crowd condition measurements. Showing a high crowd density during the event, spread arrivals before the event, fast departures after the event, and further standard deviation indicators.

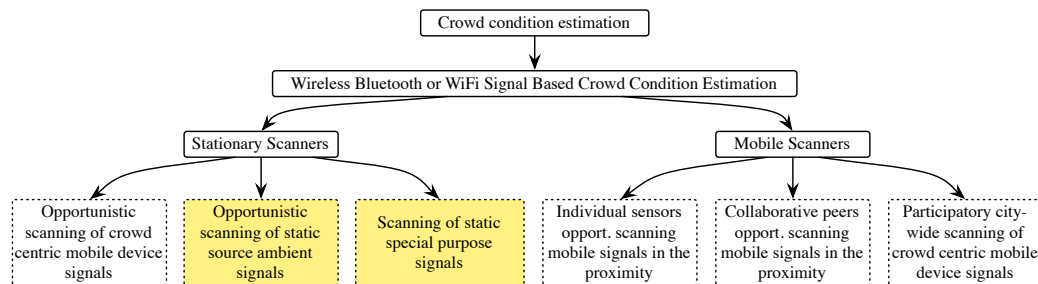
## 2.9 Conclusions

In this chapter crowd condition estimation was demonstrated with wireless opportunistic scanning of crowd devices by stationary scanners. A meshed scanner setup has been proposed and implemented in multiple large scale real-world experiments in cooperation with a large German car manufacturer. Methods in combination with supervised machine learning algorithms were presented for the crowd density estimation. The approach was presented for the proposed WiFi based aggregated crowd density estimation methods including the comparison with the reference method and the time window based concept (important for the transfer to local crowd density estimation) with varying time windows. The methods were evaluated with the comprehensive data set and ground truth annotations extracted from video based footage. Multiple methods and machine learning algorithms were compared. The aggregated crowd density estimation with the best method resulted in a low mean absolute error of just 0.0115 people per square meter. The methods were transferred to the local crowd density estimation scale, including the definition of the local crowd density scale, the Voronoi cell binning, the method, the calibration approach of the local crowd density based on the global crowd density, the calibrations coefficients, and finally presented the estimation results for local crowd density estimation with visualizations and qualitatively evaluated with visual ground truth. The approach of crowd movement estimation was demonstrated and results for local crowd movement estimation were validated qualitatively by visualizations and visual ground truth comparisons. Finally, contextual crowd condition anomaly detection methods were proposed and could be successfully demonstrated based on two significant events, the visit of the German chancellor Dr. Merkel and the large car manufacturer press conference, during in the real-world environment.

# 3

## Crowd Condition Estimation with Stationary Sensors Scanning Stationary Devices

Jens Weppner, Benjamin Bischke, and Paul Lukowicz. Sensing room occupancy levels with ieee 802.11n wifi channel state information fingerprinting. *IEEE Sensors Letter*, 2017



**Figure 3.1.** Thesis outline and wireless signal based crowd condition estimation scanning ontology.

### 3.1 Problem Statement

Knowing room occupancy levels (how many people are in a given room) is important for a number of applications. Examples range from evacuation coordination in emergency situations through the optimization of office space usage to energy management. Given the fact that today WiFi can be found in nearly all buildings and public spaces, using the disturbance that people cause, WiFi signals as a ‘virtual sensor’ is an attractive approach for occupancy level estimation. The problem to solve is to find an approach where the crowd is fully passive and no opportunistic mobile signals sources are used, which means that no re-calibration is necessary on potentially varying crowd devices in the future. However, WiFi signal analysis has mostly been used for estimation of qualitative crowd density (which is associated with strong variation in the signal strength) and for the detection of movement (which cause characteristic temporal signal fluctuations, see related work below). The estimation of the number of people in a small group ( $\leq 10$ ) who is largely static (e.g. working at their desks) in a shared space such as an office is a more difficult problem as it is related to subtle, mostly static signal changes. To address this problem we present a new sensing concept that applies machine-learning techniques to appropriate features extracted from the Channel State Information (CSI) data.

### 3.2 Chapter Overview

In this chapter the approach of analyzing signals is presented based on stationary scanners opportunistically sensing ambient WiFi access points and stationary scanners sensing special purpose signal sources. The chapter includes related work, the theoretical foundation of signal path loss and signal propagation properties, signal analyzing methods, experiments, and results.

In section 3.5 a method is proposed for estimating the number of people present within an exhibition hall by analyzing signals from over 80 opportunistic/ambient wireless access points (not under own control). When looking at phenomenas that are determined by a high degree of signal blockage (e.g. detecting a dense crowd) much of the complexity can be ignored as the RSSI can be used for analysis. A

method to estimate the fill level is presented and evaluated based on video based ground truth information.

In section 3.6 a fingerprinting method is proposed for estimating the number of people present in a room (e.g. in a shared office space) from signal-to-noise-ratio (SNR) and signal phase (PHASE) data provided by the the IEEE 802.11n CSI (Channel State Information). We apply random decision forests machine learning to SNR and PHASE based features and show that the exact number of people can be estimated with a precision of 0.67 and approximate occupancy level ranges with a precision of 0.87 at an affordable cost. We evaluate our approach in two settings: one small room ( $20 m^2$ ) with 0 to 2 and one medium ( $60 m^2$ ) office space with 0 to 8 people doing their work at desks as usual. Beyond determining maximum recognition rates we systematically investigate the impact of different design choices (antennas, features, fingerprint density) on system performance.

### 3.3 Related Work

First introduced by Woyach et al. [107] RSSI variations were empirically described for objects moving between and in the vicinity of signal source and scanner in lab experiments. They also identified that RSSI deviations at 2.4GHz are doubled compared to 433MHz signals. Related experiments were performed in office floors where the number of people where low, i.e. up to 4 subjects have been counted in a  $150m^2$  and  $500m^2$  [108] but focussing in passive localization of individuals. Other related work concerns small spaces with single WiFi access points in short distance to the people and counting or detecting presence of few people, i.e. with 16 sensors in a 7x7m area [109], three rooms spanning 9x15m area [93] in contrast to crowd condition estimation. More distant to the work presented in this chapter but also building on signal attenuation and crowd awareness is done by Hiroi et al. [41]. They proposed a RSSI crowd compensation scheme for indoor localization and demonstrated this with 40 Bluetooth low energy (BLE) scanners and 100 stationary BLE tags while 100 people were present. Their localization method considering the crowd density yielded a 59.8% higher accuracy than a simple positioning method without compensation. Related work further away is indoor localization where

known access point locations are used for trilateration/fingerprinting. In that case, access points are scanned by mobile devices.

Different sensing technologies have been proposed to estimate the occupancy level of spaces. These include surveillance cameras [25], thermal sensors [94], pressure sensors [22], acoustic sensors [49], and floor pressure sensors [69]. Compared to most of the above sensing modalities WiFi signal analysis has the advantage of requiring no extra hardware installation as WiFi is present in most spaces and just the special drivers need to be installed and being less privacy intrusive than computer vision or thermal cameras. Work so far includes strategies evaluating subjects' beacons (i.e. smart-phones) for crowd density estimation by collaborative mobile sensors [100] by citywide mobile participatory sensors [101] or by stationary sensors [102] based on device identifier and RSSI features. RSSI methods are suitable for qualitative crowd density estimation, indoor localization and many other interesting applications, but its limits arise when subtle signal variations due to few people being present need to be measured. With respect to the more general use of CSI Halperin et al. [37] published the enabling work on Channel State Information (CSI) measurement which was intended for WiFi MIMO-transmission optimizations.

### 3.4 Signal Path Loss and Signal Propagation

Wireless signal path loss occurs in free space and when signals interact with obstacles. Such obstacles can be static structures, or transient obstacles such as human bodies. Signal path loss for free space and static structures is well understood and is theoretically described with a formula depending on the distance and material the signal is propagating through. In an ideal environment the free-space signal propagation loss  $L_p$  [9] is be given by

$$L_p = 20 \log_{10}(f) + 10 n \log_{10}(d) 10^{-3} - 27.55[dB].$$

Where  $f=2400$  MHz) is the signal frequency,  $n$  is the path loss exponent ( $n_{\text{freespace}} = 2$ ) and  $d$  [meter] is the distance between the source and sensor. Phillips et al. [78] analyzed the prediction of the wireless path loss towards wire-

less signal coverage mapping methods. While signal coverage mapping methods predominantly consider static structures (walls, ceilings, columns, and furniture etc.) they also demonstrated the influence of transient obstacles. They excellently described the situation of path loss prediction:

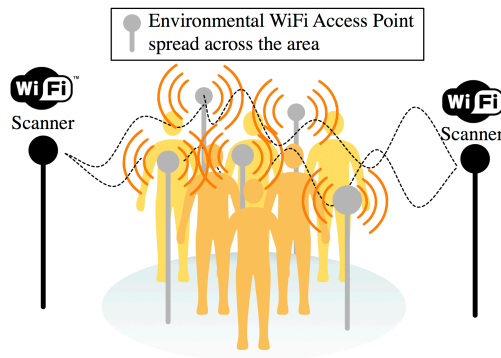
“Actually, things are a bit more complicated than this. Because an antenna radiates its signal simultaneously in all directions, the signal can take many paths to the receiver. Each path may interact with the environment in a chaotically different way and arrive at the receiver delayed by some amount. If these delayed signals are in phase with one another, then they produce constructive interference. If they are out of phase with one another, they produce destructive interference. The spread of this delay is called the delay spread and the resulting attenuation is called multi-path fading. When this attenuation is caused by large unmoving obstacles, it is referred to as shadowing, slow-fading, or large-scale fading and when it is caused by small transient obstacles, and varies with time, it is called scattering, fast fading, or small scale fading. When the signals interact with the environment, they can be delayed by reflections, or frequency-shifted by diffractions. Mobile transceivers also incur frequency shift due to Doppler spreading. Frequency shifts and delay spread both contribute to small scale fading.” [78]

Path loss and signal propagation is a key aspect in wireless crowd condition estimation with Bluetooth and WiFi scanning. Unfortunately, the modeling approach has its limitations as the location of opportunistic transceivers such as Bluetooth/WiFi smartphones is unknown and the people are moving. In most applications, the additional error of small scale signal fading is computed ‘stochastically’ using a probability distribution (often Raleigh, although Ricean and m-Nakagami are popular). Phillips et al. [78] surveyed wireless signal path loss models and described large-scale signal fading (large unmoving obstacles) and small scale signal fading (small transient obstacles varying with time). They state that exhaustive Models cannot, without perfect knowledge of the environment, be expected to predict the small-scale fast fading due to destructive interference from multi-path effects and small scatterers (which varies with time  $t$ ). Philips et al. state that general machine learning approaches in the domain of path loss modeling and coverage mapping is



currently unexplored. The work in this thesis chapter is going into the direction of the machine learning approach in the domain of signal path loss abstract modeling.

### 3.5 Opportunistic Scanning of Ambient Static Devices



**Figure 3.2.** Symbolic illustration of opportunistically scanning ambient stationary devices (not under control) with people influencing the wireless signal properties.

In this section an approach for estimating the crowd condition based on analyzing the received signal strengths of static and opportunistic common WiFi access points is presented. A method is proposed for estimating the number of people present within an exhibition hall by analyzing signals from over 80 opportunistic/ambient wireless access points (not under own control). When looking at phenomena that are determined by high degree of signal blockage (e.g. detecting a dense crowd) much of the complexity can be ignored as the RSSI can be used for analysis. The method is evaluated based on video based ground truth information. Finally, the results of the proposed method is compared to the reference method ‘visit periods’ of opportunistic crowd devices as described in the previous chapter subsection 2.8.2.

#### 3.5.1 CeBIT Experiment Context and Data Set

The scientific exploitation of the experiment goes far beyond the common scientific laboratory experiment scale. The experiment covers a large number of ambient access points (nature of a technology exhibition) and a mass of visitors. The experiment covered all days of the exhibition with different number of visitors. The

experimental environment includes simultaneous scanning of WiFi access point signal strengths. The experimental environment also includes ground truth with accurate manual human annotations extracted from the ground truth video stream. The experiments exceed previous experiments in scale, and in scientific evaluation complexity (ground truth evaluation with machine learning).

The experiment was performed during one of the largest computer exhibitions worldwide in the year of 2014. The trade fair ‘CeBIT’ is held each year on the Hannover fairground, the world’s largest fairground, in Hannover, Germany. The main audience of the exhibition are professional visitors attending during the 5 opening days of the exhibition. The exhibition attracted 210 000 visitors.

In cooperation with the Messe AG organizing committee a setup of WiFi scanners in hall 9 for the scientific experiment during the exhibition was coordinated. We performed an experiment during four sequential business days and continuously scanned the RSSI of 80 opportunistic stationary access points. The organizers and a majority of the exhibitors are setting up individual access points. We deployed 12 stationary scanners throughout the boundary of the exhibition hall (see Figure 3.3). The scanning unit is identically to the description in section 2.7. The scanners were mounted at 10 meters heights along the wall. The scan units were equipped with directional WiFi antennas directed to the area in front. Ground truth was collected with video cameras and manual annotations. We selected a thorough camera coverage setup where multiple cameras cover the whole hall. Annotations were done every 30 minutes in time for each video stream. Resulting in a total of 768 manually annotated images with up to 2500 annotated people per frame.

### **3.5.2 Methods and Results**

The approach of opportunistically scanning common wireless access points is based on the continuous signal emissions. Access points continuously send ‘beacon frames’ broadcasting the presence and information about the network. A beacon frame contains the timestamp for clock synchronization between stations, beacon interval, and network capability information such as the human readable WiFi Service set identifier (SSID). WiFi beacon frames are known as WiFi management frames,

handling roaming of clients between wireless access points and presenting the existence of the wireless network to the user.

Along with each beacon transmission the received signal strength indicator (RSSI) is obtained by the scanner unit (see section 2.7). The proposed method is based on the connection between the absolute RSSI and the number of people being present within the hall. The values are extracted by four sequential steps including signal smoothing, normalization, and the aggregation of multiple scanners. Firstly, the rolling average over a window size of 6 minutes (with a scan interval of 40 seconds this equals to 10 samples per access point) is calculated to smooth the signal. Secondly, the signal is normalized according to the reference signal strength at the empty crowd state. The empty crowd state was recognized at night (all access points were continuously enabled during the day and night time). For each pair of access points and scanners the RSSI is normalized to the value of 1.0. This step compensates different signal powers of the access points and different (unknown) distances between access point and scanner. Thirdly, all values are aggregated by each scanner by rolling window averaging the value at each point in time over all access points. Figure 3.4 gives an overview about the data processing in the first two steps.

To evaluate the relationship between the absolute RSSI and the number of people being present, the machine learning algorithm linear regression with 10-fold cross-validation was applied to the measurements described previously along with the extracted ground truth. The linear regression was trained on each unique opportunistic SSID (see step 3 in Figure 3.4). As each access point is representing the crowd influence within its own region, the resulting estimated number of people is aggregated by all 80 linear regression estimations resulting in the estimation shown in step 4 in Figure 3.4. This process is repeated for each of the four experiment days. The regression training and validation is performed for each individual day and for all days together, to identify deviation between days and to evaluate a universal estimator enabling training and validation on different days.

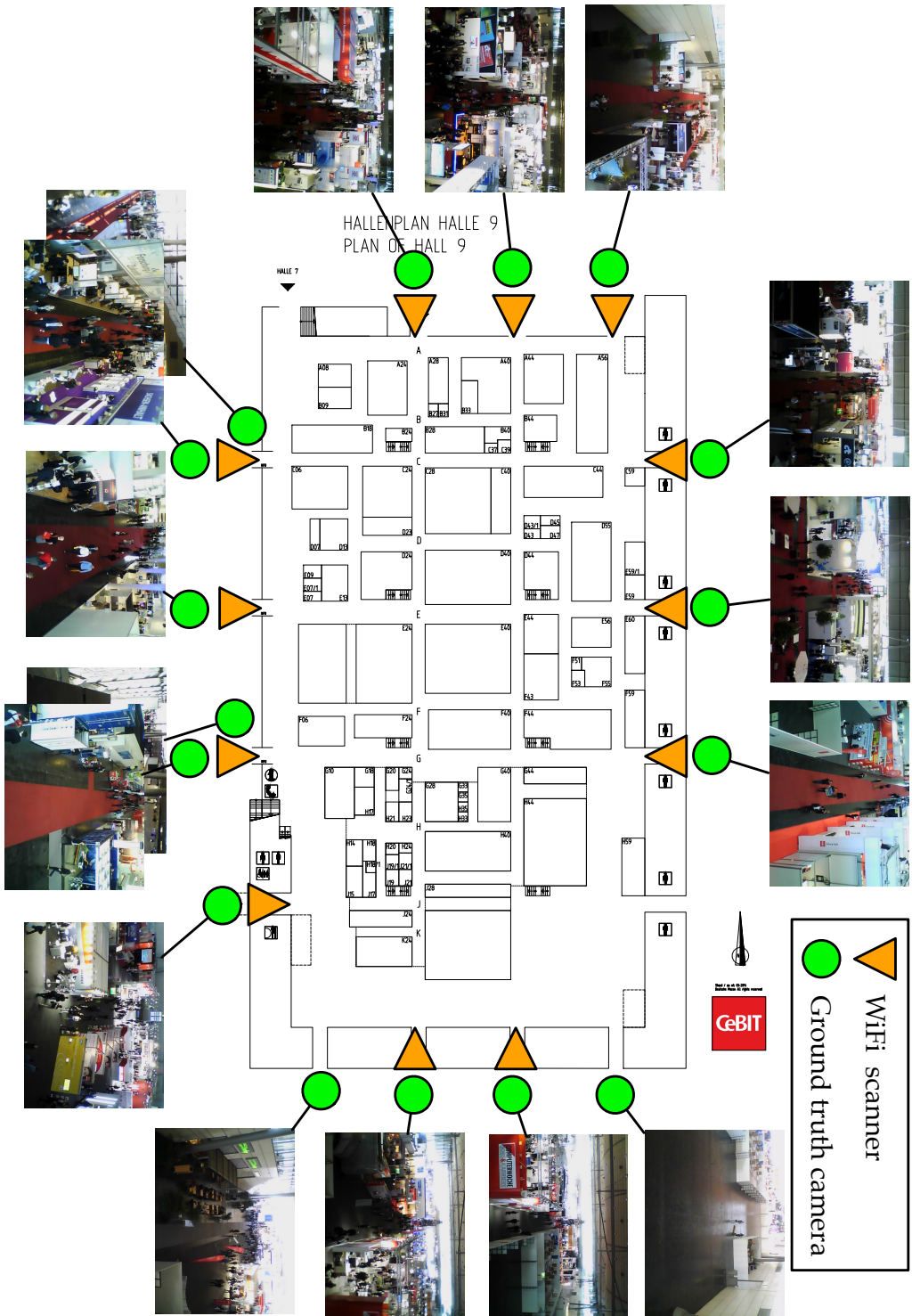
The evaluation results in a mean absolute error of 208 to 217 people when trained on individual days and 263 people when trained on all days. The mean absolute error rates have to be seen in context of in average 1700 (up to 2500) people being present.

We compared the results of this method with the reference method of scanning opportunistic crowd devices and the method described in subsection 2.8.2. For the compared method the mean absolute error is between 167 and 199 people being present. The proposed method underperforms the reference method. However, the mean absolute errors are within the same magnitude and are not significantly worse compared to the absolute number of people being present. The proposed method on the other hand is fully independent of variations of the scanned crowd devices.

**Table 3.1.** Evaluation results of estimating the number of people within the exhibition hall based on opportunistic RSSI measurements of 80 access points. The results are based on machine learning evaluation with the linear regression algorithm and 10-fold cross validation.

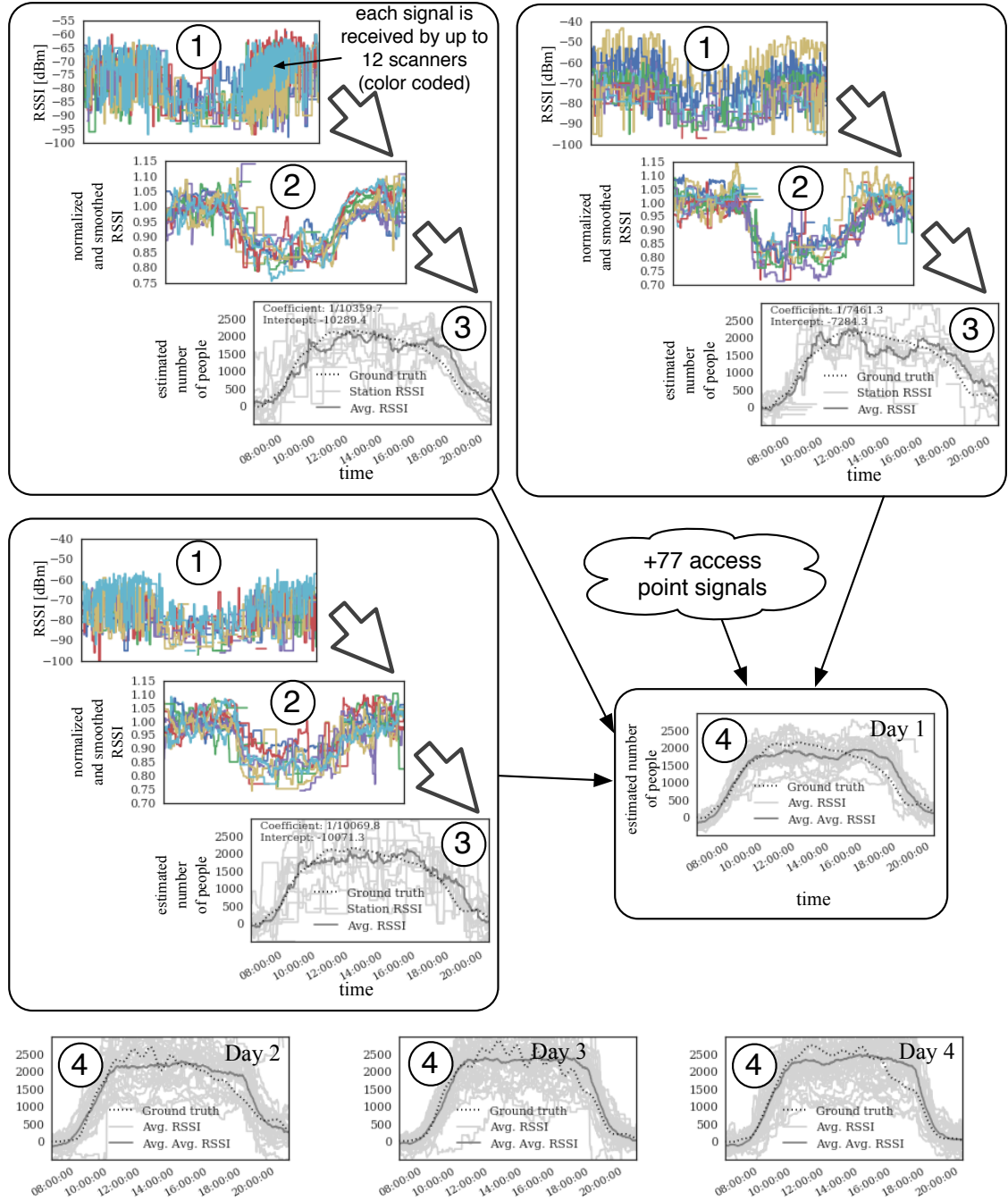
Method	Day	Ground Truth		Estimation Results	
		People Count (Mean)	People Count (Median)	Absolute Error (Mean)	Absolute Error (Median)
Ambient static access point signals	all	1696	<b>1905</b>	263	<b>207</b>
	1	1428	1669	217	213
	2	1590	1769	208	167
	3	1802	2116	210	188
	4	1945	2318	212	151
Crowd opportunistic signals	all	1696	<b>1905</b>	223	<b>190</b>
	1	1428	1669	167	134
	2	1590	1769	199	187
	3	1802	2116	187	153
	4	1945	2318	186	175

### 3 Crowd Condition Estimation with Stationary Sensors Scanning Stationary Devices



**Figure 3.3.** Setup used for evaluating the proposed method for estimating the number of people being present within an exhibition hall by analyzing signals from over 80 opportunistic/ambient wireless access points (not under own control). Technology exhibition (CeBIT in Hannover, Germany) and continuously scanned the RSSI of 80 opportunistic stationary access points. The Figure contains the locations of the 12 stationary scanners and the deployed ground truth cameras for evaluating the methods.

### 3.5 Opportunistic Scanning of Ambient Static Devices



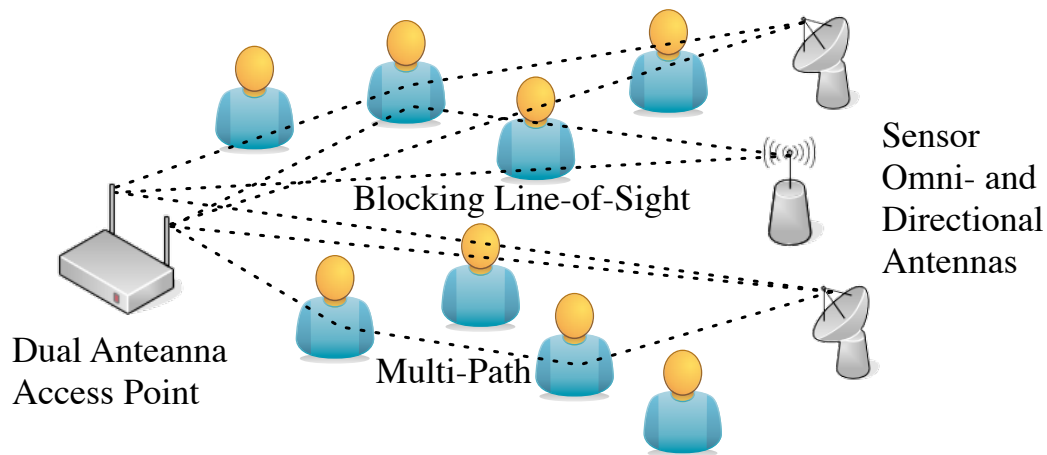
**Figure 3.4.** Data transformation process and the estimation results of number of people being present within the exhibition hall visualized as time series. The data processing and estimations contain four sequential steps including (1) RSSI signal collection by 12 WiFi scanners, (2) signal smoothing and normalization, (3) machine learning based on ground truth, and (4) the area-wide aggregation by combining 80 estimation results from each access point signal.

## 3.6 Scanning of Special Purpose Devices

In this section an approach for estimating the number of people within a small shared office space is presented. The main contributions are a method to estimate the exact number of people in an office room based on measuring WiFi Channel State Information (CSI) properties, evaluate different types of antenna setups, and determine whether a complex antenna setup is improving the estimation results compared to a single antenna setup in small scale and low number of people scenarios. The results are compared to the approach of RSSI sensing. Two experimental evaluations were performed based on a small office room and a medium-sized shared office space. The interaction of WiFi signals with the environment is a complex process that involves absorption, reflections (multi-path) and a variety of wave specific effects (refraction, interference etc.). When looking at phenomena that are determined by high degree of signal blockage (e.g. detecting a dense crowd) much of the complexity can be ignored as the received signal strength (given by RSSI) can be used for analysis. However, when considering subtle influences caused by a small number of largely static people a more complex metric is needed. In the IEEE 802.11n standard such a metric is provided by Channel State Information (CSI, [44, 37]) that captures signal strength, signal to noise ratio, and phase information for OFDM (Orthogonal Frequency Division Multiplexing) subcarriers and between each pair of transmit-receive antennas. It has originally been defined to allow the sender to improve the wireless link via transmit beam-forming [44].

### 3.6.1 DFKI Experimental Environment and Hardware

A sensing unit was built based on a INTEL NUC mini PC running Linux with a wireless Intel 5300 mini-PCI card with Multi-Input-Multiple-Output (MIMO) multi antenna support. The wireless card was selected to support a custom driver to access the signal to noise and signal phase measurements normally only used by internal WiFi management procedures [37]. The sensing unit was customized with external antenna connections specifically for the experimental setups. One directly attached external omni-directional antenna and two displaced directional antennas



**Figure 3.5.** WiFi signals are influenced by two effects: blocking of the line of sight and by multi-path effects due to objects as well as human bodies. A radio signal arrives at the receiver through different paths. The challenge is to measure and learn from such transient signal variations caused by people being present and infer the number of people.

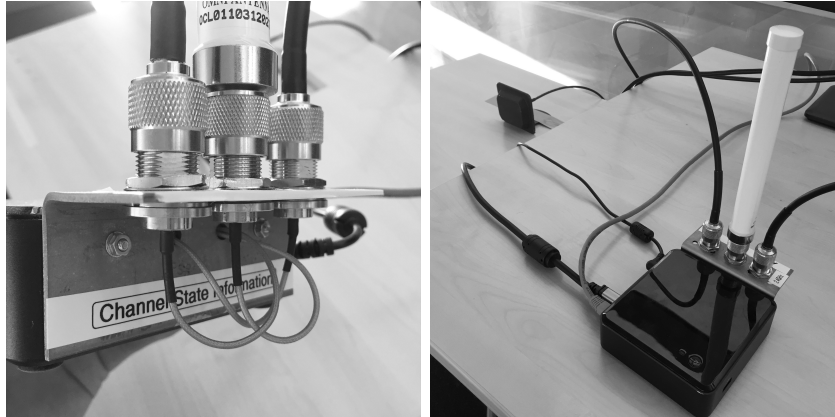
were connected (see Figure 3.6). The setup was implemented to continuously transfer data at 200 millisecond intervals. The retrieval of CSI measurements is relying on active data transmission on the wireless connection. As the special setup had the primary purpose of sensing, artificial data was transferred over the link. However, future scenarios are expected to rely on common hardware and common data transmissions.

For the small ( $20m^2$ ) and medium-sized ( $60m^2$ ) office environment the antennas were setup carefully. In the small office 0, 1, or 2 people were present at their designated desks. The sensing antennas were placed together with the directional antennas on one side of the room. The directional antennas were displaced by 1 meter to the side of the central omni-directional antenna. The directional antennas were directed towards the seat position at the given desk. A common MIMO access point served as the wireless signal emitter. The signal emitter was placed on the opposite side of the room with two displaced external antennas. The emitter antennas were placed according to the desk to be in the line of sight between signal emitter and sensor antenna.

In the medium office space 0 to 8 people were present at their desks. Four desks



were setup to face four other desks in the center of the room. The sensor antennas were placed at one side of the room. The directional sensor antennas were displaced sideways to create a line of sight through the desk seat position with the wireless emitter antennas on the opposite of the room (see illustration in Figure 3.5).



**Figure 3.6.** Sensing unit hardware for scanning a special purpose connection-oriented wireless signal source. A modified Intel ‘Next Unit of Computing’ (NUC) small form factor mini computing unit with embedded multi-antenna capable miniPCI WiFi card (Intel 5300) compatible for channel state information (CSI) scanning. CSI was originally intended for MIMO-WiFi channel monitoring for bandwidth optimization.

### 3.6.2 Methods

The proposed method is based on learning the signal properties over time. Firstly, a varying number of people being present contributes towards learning the exact count of people currently being present. Secondly, variations induced by different people and by different occupied desks is contributing towards learning the exact count of people. This approach is different to the signal fingerprinting approach in WiFi based indoor localization, but relies on creating and learning fingerprints in a changing environment and the estimation result is the number of people being present.

To model the complex process of the interaction of WiFi signals with the environment the method is extended beyond RSSI and based on the WiFi CSI measurements including signal to noise ratio (SNR) and the relative signal phase

difference (PHASE). The SNR is reflecting the relationship between the signal and the noise. As the signal power is continuous and the setup of the antenna is static, the impact of the measurement is due to noise induced to signal obstructing objects in the environment. As objects such as furniture and computer hardware is static, the change in signal noise is due to human bodies. Wireless signal noise is also induced by other wireless signals. However, within multiple weeks of the experiment such variations are assumed to be learned by the classifier to not be connected to the number of people currently being present. Similar argumentation is valid for PHASE measurements. The SNR is extracted for each pair of sender and sensor antenna (in total six antenna combinations). The absolute wireless signal phase measurement is infeasible as the sender and receiver must be perfectly synchronized. Unfortunately, commercial WiFi devices have non-negligible carrier frequency offsets. Nevertheless, we can identify and calculate the signal phase difference by the two signal streams emitted by both sender antennas synchronously. The signal phase difference is extracted for each sensor antenna (two measurements). The method relies on the time windowing approach with a time window size of one second to allow real-time estimations even when fast fluctuations in the number of people occurs. The method calculates absolute and relative signal properties reflecting the long term (absolute) and short term (variations) of the signal. Based on the raw SNR measurement the proposed method is calculating the properties of sequential measurements within one window. We define the properties as the mean SNR value, the variance SNR value, the minimum SNR and maximum SNR value within a one second window. Similarly, based on the PHASE the signal properties are defined as the PHASE mean, the PHASE variance, the PHASE minimum, and the PHASE maximum within a one second window. The signal property computation is repeated for each of the 30 accessible WiFi signal sub-carriers (base frequencies within the frequency band). Multiple frequencies with small variations increase the feature dimensions and allow the classifier to learn from small variations based on different frequencies at the exact same time.

To learn fingerprints of multiple extracted properties from the signals the machine learning classification principle is selected as no direct relationship between the fingerprints is expected to be suitable for a regression approach. The machine learning ensemble algorithm ‘random decision forests’ was selected being able to

handle multiple dimensions and a strong variation in machine learning features very well. The machine learning algorithm was trained and validated with the common 10-fold cross-validation. As stated in the previous chapter, collecting ground truth is a crucial step for supervised machine learning and to thoroughly evaluate the classification estimates. The raw ground truth was collected by a video camera covering the room. The video stream was then manually annotated by students at each 5 seconds of the video stream with the exact position of each person and the derived true number of people.

The following sub-section describes the detailed application of the machine learning process towards answering the research questions.

### 3.6.3 Experimental Validation and Aims

Besides estimating the number of people being present within an office, the experimental validation includes multiple aims. The aims are briefly described in the following list.

- To evaluate the surplus value of the CSI measurements the proposed method is compared to the previously presented method based on RSSI measurements towards estimating the exact number of people being present within a room.
- A comparison is performed between a pure statistical knowledge about the number of people being present at different times (basic reference method) and the estimated number of people being present with the proposed method.
- Different configurations of the antenna setups are compared to evaluate the advantage of a complex antenna setup over a simpler setup. This has the aim to assess the cost of the setup compared to the estimation result. A simple setup has the advantage to be implemented into future common access points without the need for specific external antenna setups according to the line of sight situation of people being present. One scenario includes the evaluation of a single omni-directional antenna, similar to an antenna in common access points. Another scenario includes the evaluation of two carefully installed directional antennas targeted towards the expected locations (desk rows) of people being present. Another scenario includes the evaluation of all three antennas including one omni-directional plus two carefully positioned directional antennas.

- The importance of individual signal measurements and the applied methods are evaluated. The evaluation is based on using a sub-set of signal measurements including the evaluation of different antenna scenarios. The signal measurements SNR and PHASE are evaluated in connection with the time windowing methods of the mean value, the standard deviation value, the minimum value, and the maximum value.
- Depending on the application either an approximate estimation of the precise number of people is required or a precise estimation of a range of people. Towards comparing both scenarios the people count estimation is evaluated on the precise number of 0, 1, 2, 3, 4, 5, 6, 7, or 8 persons and on the range of 0, 1–3, or 4–8 persons being present.
- It is assumed that the machine learning classifier training phase needs to learn different permutations of people being present in the room. The impact of relying only on a sub-set of permutations for training is analyzed. Ascending fractions of the training data (‘fingerprints’) is selected for classifier training.

The experimental validations are applied to the small office space and to the medium office space environment, while the focus is on RSSI/CSI comparison in the small office scenario and the focus is on the other validation aims in the medium office scenario. Due to the low number of permutations of people in the two person office scenario a detailed evaluation is omitted. The used classification metrics are the precision, recall, and f1-score (harmonic mean between precision and recall) defined as follows, where ‘tp’ is the number of true positive estimations and ‘fp’ is the number of false positives:

$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{Recall} = \frac{tp}{tp + fn}$$

$$\text{F1-Score} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

## Small-Sized Office Space

The small office space scenario ( $20m^2$ ) consisted of 0, 1, or 2 people being intermittently present at their designated desks. The office routine was not interrupted during the experiment but relied on different presence times of employees during the day over two weeks. Due to different meetings and business trips the presence varied during the experiment and different numbers of people could be detected. The scenario was evaluated towards the validation of the surplus value of the CSI based approach compared to the RSSI based approach (see section 3.5). As described before the machine learning classifier ‘random decision forests’ was applied to the set of CSI signal properties including all SNR and PHASE properties. Another classifier was cross-validated with RSSI signal properties. The evaluation of the proposed CSI approach results in a f1-score of 0.73 with equal precision and recall (see Table 3.2). The evaluation of the traditional RSSI approach results in a inferior f1-score of 0.50 and near equal precision and recall (see Table 3.2).

**Table 3.2.** Small office experiment results. SNR&PHASE scanner in comparison with traditional RSSI scanner. Classification between 0, 1 or 2 persons in room.

Antennas	Features	Classification results		
		F1-Score	Precision	Recall
ALL	RSSI	0.50	0.50	0.51
ALL	SNR&PHASE	0.73	0.73	0.73

## Medium-Sized Office Space

The medium-sized office space environment ( $60m^2$ ) consisted of 0, 1, 2, 3, 4, 5, 6, 7, or 8 people being intermittently present at their designated desks. The office routine was not interrupted during the experiment but relied on different presence times of employees during the day over two weeks. The shared office space is a room where complementary science employees work at different weekdays and during different hours of the day. Due to a large number of complementary science

employees their presences varied during the experiment and different numbers of people could be detected. If not otherwise declared the following validation is based on the exact estimation of 0,1,2,3,4,5,6,7 or 8 people being present.

To evaluate the surplus value of the CSI measurements the proposed method is compared to the previously presented method based on RSSI measurements towards estimating the exact number of people being present within a room. A statistical reference based on hourly knowledge of the occupation of the room is extracted from the ground truth to be compared with the machine learning classifier. All SNR and PHASE properties are extracted and cross-validated together with the ground truth by the random decision tree forest classifier. The statistical reference method resulted in a weak f1-score of 0.27 (with similar precision and recall). The CSI approach has a f1-score of 0.67 (see Table 3.3). This means that the classifier is superior to the statistical reference method, but still involves mis-interpretations in the true count of people being present.

Different configurations of the antenna setups are compared to evaluate the advantage of a complex antenna setup over a simpler setup. Firstly, the scenario with a single omni-directional antenna was evaluated. The random decision tree forest classifier was cross-validated with all extracted signal properties and resulted in a f1-score of 0.61 (see Table 3.4). Secondly, the scenario with two directional antennas was evaluated. The random decision tree forest classifier was cross-validated with all extracted signal properties based on the given antenna and resulted in a f1-score of 0.62. Thirdly, the scenario with three antennas (one omni-directional plus two directional antennas) was evaluated. The random decision tree forest classifier was cross-validated with all extracted signal properties based on the three antennas and resulted in a f1-score of 0.67. Comparing the three antenna configurations a clear benefit of using different antennas can be seen, while the two directional antennas result in no significant precision enhancement compared to the single directional antenna. This clearly denotes that the signal variations are not just relying in line of sight obstructions but on signal multi-path signal interferences also detected with a simple single antenna setup.

The importance of the individual signal properties is evaluated with the aim of identifying the most important properties to the classifier. The signal measurements SNR and PHASE are evaluated in connection with the time windowing methods

of the mean value, the standard deviation value, the minimum value, and the maximum value. The evaluation is based on the three antenna scenario. An individual random decision tree forest classifier was cross-validated for each mean SNR, maximum SNR, minimum SNR, and standard deviation SNR. The best random decision tree forest classifier estimation f1-score 0.67 was detected alone for each mean SNR and maximum SNR properties (see Table 3.5).

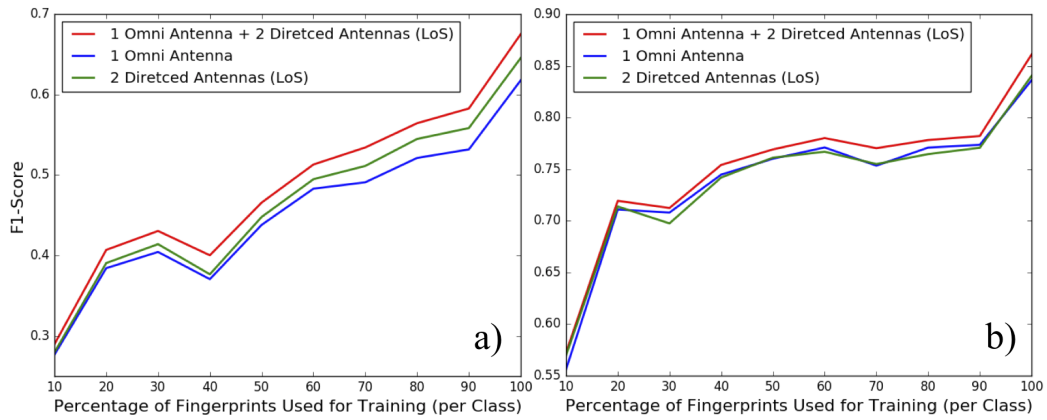
Depending on the application an approximate estimation of the precise number of people is required or a precise estimation of a range of people is required. Towards comparing both scenarios the people count estimation is evaluated on the precise number of 0, 1, 2, 3, 4, 5, 6, 7, or 8 persons and on the range of 0, 1–3, or 4–8 persons being present. While the random decision tree forest classifier based on the exact people count resulted in a f1-score of 0.67, the ranged people count resulted in a f1-score of 0.87 (see Table 3.6). This demonstrated that the precision and recall can be significantly improved on a ranged estimation of people.

It was assumed that the machine learning classifier training phase needs to learn a significant amount of different permutations of people being present at different locations in the room. The impact of relying only on a sub-set of permutations for training is analyzed. Each sub-set of permutations includes each occurring class (0–8 people) but for each class only a fraction of all occurring seat permutations (fingerprints) are included. The number of permutations is defined by the binomial coefficient  $\binom{n}{k}$ , where  $n = 8$  is the number of desk seat locations and  $k$  is the number of people being present. For example class ‘0’ includes only one permutation: the empty permutation. Class ‘1’ includes 8 permutations on all seats. Class ‘2’ includes 28 permutations, class ‘3’ includes 56 permutations and so on. Based on these permutations multiple fractions are selected in 10% steps from 10% to 100% of each of the class permutations. One random decision tree forest classifier was cross-validated for each of the selected sub-sets based on the three antennas. In Figure 3.7 the ordered sequence of classifications with sub-sets is presented for the exact estimation of the number of people and the ranged estimation of the number of people. It can be observed that with an increasing fraction of permutations the estimation f1-score can be improved continuously. However, an increased number of fractions used for classifier training requires an increased number of ground truth labels is necessary for supervised learning which raises the complexity of training

the classifier in a given environment.

**Table 3.3.** Medium office space experiment results on baseline statistical classifier (average hourly ground truth statistics). Ground truth is aggregated to hourly value and classification based on overall learned hourly values. The Table presents the inferior results of a simple hourly statistical evaluation which is due to fluctuations of office space fill level over days.

Approach	Classification Results		
	F1-Score	Precision	Recall
Baseline statistics	0.26	0.27	0.26
SNR&PHASE (all antennas)	0.67	0.66	0.66



**Figure 3.7.** Medium office space experiment results. Influence of the percentage of available fingerprints used for training the classifier. a) Classification between 0,1,2,3,4,5,6,7,8 persons. b) Classification on 0,1–3 and 4–8 persons. The Figure shows the importance of learning sensor data not just at different fill levels but also on different positions of people.



**Table 3.4.** Medium office space experiment results. Comparison between different antenna setups and signal characteristics: 1 omni-directional antenna, 2 directional antennas, or 1 omni-directional + 2 directional antennas. Using multiple features on signal characteristics of signal-to-noise (SNR) and signal phase (PHASE) measurements. Classification between 0,1,2,3,4,5,6,7 and 8 persons in shared office space.

Antennas	Feature Sets	Classification results		
		F1-Score	Precision	Recall
OMNI	SNR	0.59	0.59	0.59
	PHASE	0.51	0.52	0.52
	SNR&PHASE	0.61	0.61	0.60
DIRECTIONAL	SNR	0.61	0.61	0.61
	PHASE	0.52	0.53	0.52
	SNR&PHASE	0.62	0.62	0.62
ALL	SNR	0.65	0.65	0.65
	PHASE	0.58	0.58	0.57
	SNR&PHASE	0.67	0.66	0.66

**Table 3.5.** Medium office space experiment results. Comparison between different feature methods with signal-to-noise-ratio (SNR) and signal phase (PHASE) with 1 omni-directional + 2 directional antennas. Classification between 0,1,2,3,4,5,6,7 and 8 persons in shared office space.

Feature	Method	Classification results		
		F1-Score	Precision	Recall
SNR	MEAN	0.67	0.67	0.67
	STD	0.43	0.44	0.44
	MIN	0.65	0.65	0.65
	MAX	0.67	0.67	0.67
PHASE	MEAN	0.55	0.56	0.55
	STD	0.33	0.34	0.34
	MIN	0.58	0.58	0.58
	MAX	0.58	0.58	0.58

**Table 3.6.** Medium office space experiment results. Comparison to classification of the ranges 0,1–3 and 4–8 persons in the shared office space.

Class Range	Classification results		
	F1-Score	Precision	Recall
0,1,2,3,4,5,6,7,8	0.67	0.66	0.66
0,1–3,4–8	0.87	0.87	0.87

### 3.7 Conclusions

In this section two scenarios were presented: Estimating the fill level of thousands of people in a large-scale exhibition hall by relying only on ambient (not under own control) static wireless access point signals and estimating the number of people (up to 8) in a small-scale office environment with special purpose wireless devices and multiple carefully setup antenna configurations.

An experiment of this scale has been first presented by the work in this chapter. We conclude that crowd condition estimation by stationary scanning of opportunistic WiFi access point signals is feasible to a limited extent. This method relies on the availability of sufficient access points and the distribution across the whole area. However, estimation errors could be detected at the state when people were rushing out of the exhibition at the end of the business day and causing signal fluctuation expected by a higher fill level. The proposed method was compared to the method in subsection 2.8.2. Compared to this, the estimation results are slightly worse, but do not rely at all on visitor device detections and can be deployed independently of the visitor backgrounds.

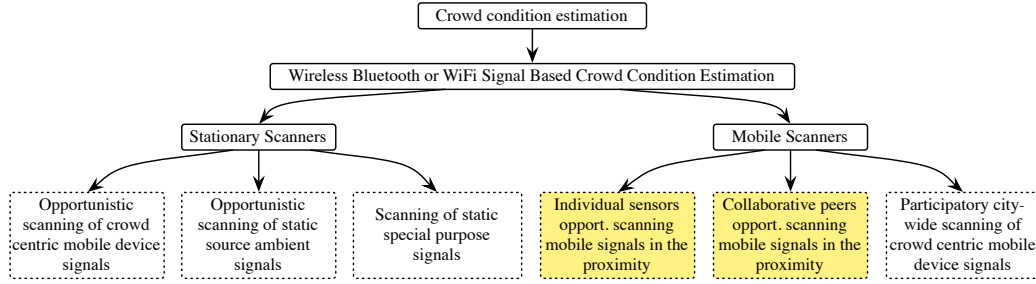
In the second part of this chapter a method was proposed to count the number of people in shared office spaces. Both the RSSI approach and the reference method of statistical knowledge were outperformed. While a simpler setup of a single omnidirectional antenna resulted in similar results as two carefully setup directional antennas, the combination of the three antennas resulted in an increased precision of 0.67. While the person precise people count estimation is the ultimate goal, currently, the estimation on a people count range is more reliable with a precision of 0.87. However, the current effort of training ‘fingerprints’ needed for different people arrangements is big. Current work could be extended in future work by minimizing the training effort. Future work could also focus on methods assessing the required amount of training data depending on the location complexity and number of people.

# 4

## Crowd Condition Estimation with Mobile Scanners Opportunistically Scanning Crowd Devices

Jens Weppner and Paul Lukowicz. Collaborative crowd density estimation with mobile phones. In *Second International Workshop on Sensing Applications on Mobile Phones. ACM Conference on Embedded Network Sensor Systems (SenSys-11)*, 9th, November 1, Seattle, USA. Microsoft, ACM, 2011

Jens Weppner and Paul Lukowicz. Bluetooth based collaborative crowd density estimation with mobile phones. In *Proceedings of the Eleventh Annual IEEE International Conference on Pervasive Computing and Communications (Percom 2013)*, pages 193–200. IEEE, 3 2013

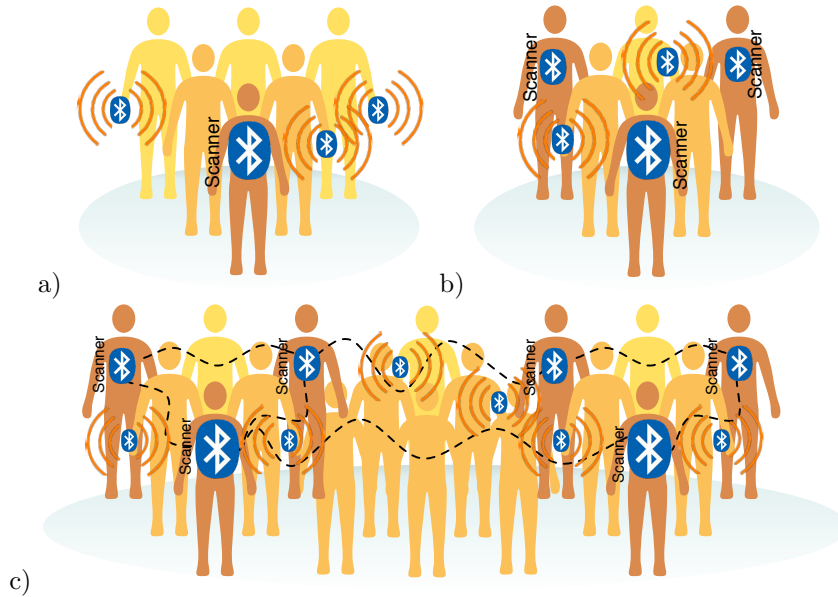


**Figure 4.1.** Thesis outline and wireless signal based crowd condition estimation scanning ontology.

## 4.1 Introduction

Knowing the density of a crowd can be relevant for a number of applications. Examples range from crowd control and emergency services through urban planning to consumer applications recommending where to go out (based where many other people also have gone to). While in some applications dedicated infrastructure such as stationary wireless crowd scanners, access control gates, or CCTV cameras [88] may be used, in others it would be desirable to be able to estimate crowd density without pre-installed infrastructure. One possibility is to recruit enough users to be able to estimate the density from the number of devices which report being in the relevant area. The obvious disadvantage of this method is that a significant number of users must be recruited, which is not always possible. In this chapter we present an alternative method that requires only few users moving through the environment with their mobiles scanning for discoverable Bluetooth devices. Since the rise of mobile phones and the advent of smartphones in 2007 the community is putting efforts to infer social state by monitoring ubiquitous wireless user centric signals. Using wireless user centric signals for crowd density estimation has gained interest since 2010. As of today we are still witnessing an explosion in ubiquitous wireless enabled devices. The environment is full of mobiles' signals from smartphones, smartwatches, or fitness-trackers. This chapter builds on previous work directed to using Bluetooth scans to analyze social context and extends it with more advanced features, leveraging combined scans between numerous mobiles, and the use of relative features that do not directly depend on the absolute number of devices in

the environment (which may vary from venue to venue).



**Figure 4.2.** Symbolic visualization of mobile individual scanning of opportunistic crowd devices and group scanning of opportunistic crowd-centric devices.

## 4.2 Chapter Overview and Contributions

This chapter presents the crowd condition estimation approach with mobile sensors scanning opportunistically crowd devices (see Figure 4.1 for the context of this chapter within the thesis). This chapter includes the approach, the related work, the general considerations, the individual scanning methods, the group formation methods, the collaborative scanning methods, and evaluations based on multiple extensive data sets.

In section 4.6 we present a Bluetooth scan based method that can detect different discrete crowd densities. The main contributions beyond the related work are as follows.

1. We rely not just on the number of devices seen by a scan, but also take into account information about average observed signal strength and the variance in both the signal strength and the number of devices. This makes the system

more robust against variations in the number of discoverable devices that may result from the background of the people in the crowd rather than the crowd density.

2. We investigate the benefit of combining the information from several devices carried by different close by users, rather than on an individual scanner.

We evaluate the method on a data set recorded during three days at the famous Munich Oktoberfest festival which is attended by hundreds of thousands of visitors from all over the world each day. Looking at four discrete densities that cover the range from a loosely occupied space (around 0.1 people per square meter) to dense crowd (around 0.4 people per square meter) we demonstrate recognition rates of 66% using both relative and absolute features. This is over 32% better than the simple approach from previous work that relies on the number of devices found only.

In section 4.7 we introduce new collaborative concept of multiple teams walking intermittently nearby and scanning each other in addition to the previous method. The main contributions beyond the related work above are as follows:

1. We rely not just on the number of devices seen by a Bluetooth scan, but also take information about the link structure between actively scanning Bluetooth devices, ratio of discovered devices in the current scan window to previous scan windows, teamwise diversity of discovered devices, number of semi-continuous device visibility periods, and device visibility durations into account.
2. We propose a method to combine the collaborative sensor information from several mobile phones carried by different groups of static and dynamic intermittently close by users (only 0.2% of all people are equipped with a Bluetooth scanning mobile phone) to determine the crowd density in an area of  $2500 m^2$ .

We evaluate the method on a data set recorded during three days at the European soccer championship public viewing event in Kaiserslautern which is attended by thousands of visitors. Looking at seven discrete densities that cover the range from a nearly empty space (around 0.01 people per  $m^2$ ) to dense crowd (above 2.0 people per  $m^2$ ) we demonstrate recognition rates of over 75% using both relative

and absolute features. This is over 30% better than the simple approach from previous work that solely relies on the number of devices found.

### 4.3 Related Work

The work most similar to this chapter is done by Nicolai et al. [75] where the discovery time of Bluetooth devices as well as the relation between number of people and number of discoverable Bluetooth devices was investigated. As opposed to our approach the work relied on static Bluetooth sensing locations and only the absolute number of discovered Bluetooth devices was used. Along the same lines Morrison et al. [70] investigated crowd density estimation in stadium-based sporting events. However, they did not attempt rigorous automatic classification and focused on a visualization tool for Bluetooth logs. Another use case of Bluetooth scanning is described in [52] by Kostakos et al. They recorded passenger journeys in public transportation by analyzing Bluetooth fingerprints. In [76] O’Neill et al. presented initial findings in Bluetooth presence and Bluetooth naming practices. Slightly further away from our work, Eagle et al. showed [30] how to recognize social patterns in daily user activity, infer relationships and identify socially significant locations from using Bluetooth scans. BLIP Systems [12] exploited a stationary Bluetooth based people tracking system. Based on multiple Bluetooth zones scenarios like queue length at airports or travel times by car are indicated. Table 4.1 presents an overview of different existing crowd condition estimation approaches with Bluetooth scanning and declares the distinction to other work.

Campbell et al. [16] and Burke et al.[15] introduced the general concept of people-centric sensing. Wirz et al. [105] demonstrated the specific need for detecting potentially critical crowd situations at an early stage during citywide mass gatherings. They collected GPS traces to create a crowd condition visualization which was monitored by the city police.

Related work so far focussed on *specific* situations for crowd condition estimation such as queues [53], crowd flow at entrance gates [76], or crowd transitions between check-points [92]. Related work introduced signal strength and signal strength variance effects [72] for an increasing number of people (0–20) between the link with passive stationary sensors and stationary transmitter in a controlled lab



environment. However, related work lacks a detailed description and thorough experimental studies at non-specific scanner locations within the crowd. Differently to previous methods we introduced an approach of estimating the crowd density at arbitrary locations.

**Table 4.1.** WiFi and Bluetooth crowd condition estimation concepts in related work requiring specific environments such as a crowd passage (i.e. queue, gate, turnstile, etc) and potentially inferring to an area-wide crowd condition. Followed by the approach we propose.

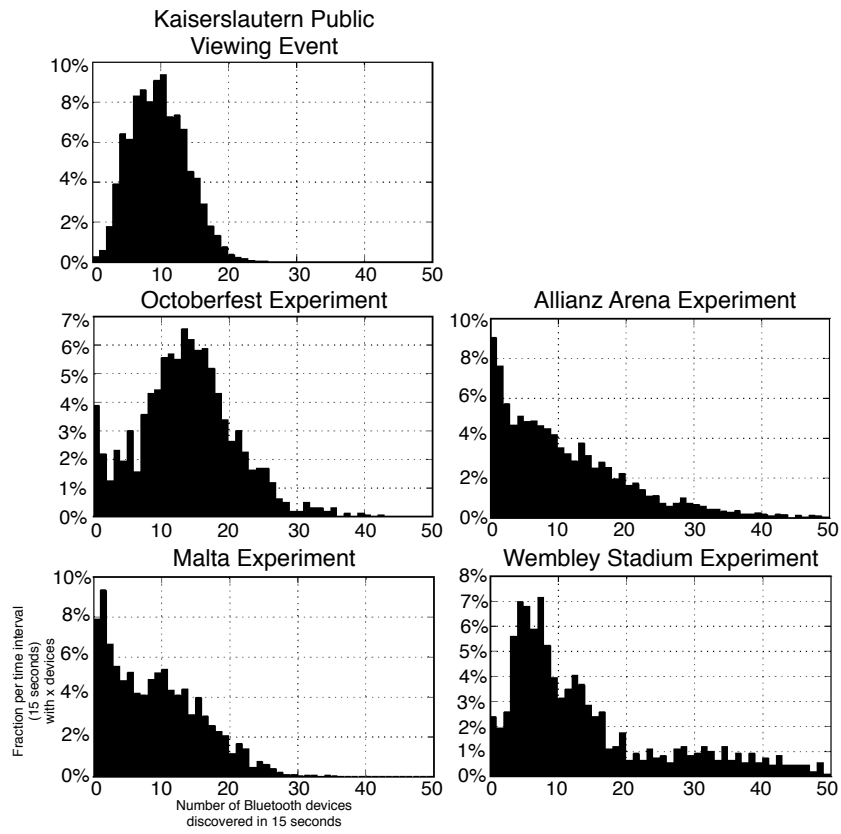
<b>Bluetooth Measurement Setup</b>	<b>Description</b>
Gate flow measurements (cumulative people counting)	A Bluetooth scanning device is positioned stationary near a narrow entry/gate/turnstile/security check point etc. This is mostly accomplished with fixed Bluetooth scanning hardware. The number of discovered Bluetooth devices is summed up over time to make a statement about the current crowd density in a bounded area. Instantaneous crowd density evaluations are not possible. [76]
Queue measurements (waiting time)	A Bluetooth scanning device is positioned near a queue or waiting area (i.e. supermarket check-out/airport check-in/public transport etc.) to measure the time between the appearance and disappearance of unique Bluetooth devices to estimate the current waiting time. This is useful for estimating a relatively small group of locally bounded stationary people. This approach does not consider dynamic people who are not bound to a single location. [53, 12]
Checkpoint measurements (people flow detection). Transition time tracking (i.e. at exhibitions, airports, cities).	Two or more specific Bluetooth scanning devices are fixed at two or more separate places with a well known distance to each other. Time is measured between a discovery of a unique device at place A and place B. This approach is similar to queue length estimation but works in a more widespread area. Requirement of this approach is that people are walking from place A to place B and that the crowd density has a direct relationship with the time needed between both places. Versichele et al. presented such an approach during a citywide festival. [92]
Solitary stationary measurements (instantaneous people counting)	One Bluetooth device is located at a specific location (i.e. shop etc.) with limited dimensions. The instantaneous number of discoverable Bluetooth devices in the covered area is mapped to a number of people by assuming a fixed proportion between the number of discoverable Bluetooth devices and the number of people. [75]
Collaborative and in combination of stationary and dynamic measurements (instantaneous in-crowd people counting)	We propose the method of multiple people equipped with ubiquitous smartphones and we present new features to evaluate the crowd-density instantaneously. The collaboration allows a coverage of a larger area and a crowd-density estimation with features which are even independent of the proportion between discoverable Bluetooth devices and the number of people. <i>This thesis chapter is based on this approach.</i>

## 4.4 Approach

The foundation of our Bluetooth based crowd density sensing technique is the general observation that many people have the Bluetooth transceivers of their mobile phone in the discoverable mode as default setting. This is illustrated in Table 4.2 and Figure 4.3 on data sets from 5 different locations and venues across Europe: (1) several soccer games from the German first and second division, collected in and around the stadium, (2) the world-famous Munich Oktoberfest festival, (3) the England-France soccer game at Wembley Stadium in November 2010, (4) a music festival in Malta, and (5) the 3-day European championship public viewing event in Kaiserslautern. We observed that most discoverable Bluetooth devices are personal smartphones and cell phones mostly manufactured by Samsung, Nokia and Sony Ericsson (see table 4.3).

From the above only the Munich Oktoberfest festival and the public viewing event in Kaiserslautern data was collected explicitly for crowd density estimation and thus contains crowd density ground truth that is used for the quantitative evaluation later in this chapter. During the Munich Oktoberfest experiment we only had a small number of people walking synchronously back and forth on the event's main street. Regarding the Oktoberfest experiment we only can utilize a subset of the features presented in this paper because of the lack of information of the bi-directional link structure between actively scanning Bluetooth devices. The Kaiserslautern public viewing experiment gives us a complex data set with asynchronously walking or standing people and all feature calculation requirements to demonstrate our approach.

The other data sets were collected for different purposes, such as inertial navigation and activity recognition. However, all data sets include regular Bluetooth scans collected over periods of days by several volunteers walking through the area of the specific event during times of different crowd density. It can be seen that the median of the number of devices discovered per scan is between 8 and 13 with thousands of distinct devices having been recognized over the course of each experiment. Figure 4.3 shows that only less than only less than 10% of the scans returned no discoverable devices and up to 50 devices were seen when in dense crowd.



**Figure 4.3.** Distribution showing the fraction of the number of Bluetooth devices discovered in a 15 second time window at multiple experiment venues.

**Table 4.2.** Statistics overview about performed Bluetooth crowd condition experiments.

Event	Duration	Participants	Number of Bluetooth scans	Average devices per scan	Median devices per scan	Unique devices
Kaiserslautern (DE) Public viewing event	3 days	10,10,10	4100	5.84	6	410
Munich (DE) Oktoberfest	3 days	3,3,3	2775	13.35	13	4454
Malta (MT) Open-air festival	3 days	12,12,12	5500	8.70	8	1088
London (UK) Wembley Stadium	1 day	6	4958	15.44	10	2509
Munich (DE) Allianz Soccer Arena	4 days	10,16,6,12	14087	10.87	8	3944

**Table 4.3.** Types of discovered Bluetooth devices

Bluetooth major device class	Fraction	Bluetooth device manufacturer	Fraction
Smartphone	72.0 %	Samsung	29 %
Mobile phone	28.0 %	Nokia	32 %
Laptop	0.2 %	Sony Ericsson	12 %
Cordless phone	0.02 %	RIM	7 %
Audio headset	0.01 %	LG	7 %
Other	0.04 %	Texas Instruments	3 %
		HTC	1 %
		Other	8 %

## 4.5 General Considerations

An obvious way to estimate crowd density is to perform a scan for discoverable devices and assume that the number of devices it returns is an indication of the number of people in the vicinity defined by Bluetooth range (typically around 10 m). Unfortunately, this simple approach contains a number of problems.

Firstly, there is the issue of sufficient statistics. With the scan limited to a radius of about 10m (approximately a circle with  $300\text{ m}^2$  area) anything between a few and a few hundred people can be within range. While in a dense crowd with a few hundred people we may get a representative sample, in less crowded areas we are likely to see very strong variations between samples. Assuming the probability of any single user having a discoverable Bluetooth device to be 10%, the probability that no device is seen when 20 people are within range is  $0.9^{20} = 0.12$ . Thus we may sometimes be in a group of people which do not even have activated mobile phones while at other times we may be surrounded by a group where everyone has an active Bluetooth device.

Secondly, there is the question of signal attenuation. At 2.4 GHz (which is the transmission frequency of Bluetooth) the human body has a high absorption coefficient. This means that in a dense crowd (where we would expect to have good statistics) the effective scan range is reduced and therefore ‘falsifying’ the results.

Finally, we have to consider cultural factors. This means that the average number of people carrying a discoverable Bluetooth device may significantly vary depending on who the persons in the crowd are. For the same crowd density at a student party of a technical university a different number of devices may be present than at a fifth division soccer game in a poor rural area.

To mitigate the influences above our method does not rely solely on the absolute number of discovered devices. Instead we also use the average signal strength and signal strength variations. In addition, we look at collaborative estimation from several (up to around 10) devices. In doing so, we focus on differential features that are not directly dependent on the absolute number of discoverable devices in the environment or the absolute signal strength. As shown in the evaluation subsection 4.6.6 and subsection 4.7.3 the above measures lead to around 30%

improvement in recognition rate over a method based on the absolute number of discovered devices.

## 4.6 Oktoberfest Experiment

### 4.6.1 Individual Scanner Feature Vector

Individual scanning is relying on an individual person roaming through the area. The person is equipped with a smartphone and software for automated discovering of Bluetooth devices. Due to the fact of low signal energy used for Bluetooth, devices are known to be in the immediate proximity.

The instantaneous area covered by an individual scanners is  $314m^2$ .

Based on the raw data we define feature vectors and later extract them from the experimental data set. A feature vector  $\bar{x}$  of the dimension  $1-k$  is defined as the composition of the input variables  $x_1...x_k$  computed for each time interval  $i$  between  $t_{i-1}$  and  $t_i$ . The time interval  $i$  is defined as the duration of one scanning interval. Each data record at  $i$  is defined as the form  $(x, Y) = (x_1...x_k, Y)$ .  $Y$  is the target variable extracted from ground truth at interval  $i$ . We present the detailed description of features  $x$  in subsection 4.6.1 and 4.6.2 which are then evaluated by the machine learning method described in subsection 4.6.3.

### Count of unique crowd devices in the proximity within a scan interval (absolute feature)

The first feature (*count of distinct Bluetooth devices*) is the most obvious feature extracted from the raw data by counting the number of different devices per scan interval. As described above the discovered devices per 60 seconds time window are the set of actual surrounding devices ( $count(i) = |\cup_{s \in i} |$  where  $s$  is a time of occurrence and  $i$  is the current interval) in this time window independent to real duration of occurrence of a device.

### Mean signal strengths (relative feature)

The second feature (*mean signal strength*) averages the signal strength of all devices in a scan interval ( $average_{signal}(i) = \frac{\sum_{s \in i} signal(s)}{count(i)}$ ), motivated by the assumption that the average signal strength gives a hint for the crowd density. Assuming surrounding people are shielding the signal the average of the signal strength might be lower. Due to the opportunistic scanning approach of crowd devices their exact location is unknown. It is unknown to the algorithm whether a low signal strength corresponds to a device being further away during a less crowded time or being close by but strongly attenuated by the crowd.

### Variance of the signal strengths (relative feature)

The third feature (*variance of the signal strengths*) is defined by the variance of the measured signal strengths of unique devices in a scan interval ( $variance_{signal}(i) = \frac{\sum_{s \in i} signal_s^2}{number_i} - average(i)^2$ ). Assuming a shielding effect is measurable, the Bluetooth signal reception is excellent from people walking nearby and the Bluetooth signal is near the reception threshold from people walking rather further afar. At high crowd densities the signal strength variance would be higher than at low crowd densities where less people are shielding the Bluetooth signal.

### Compound features (individual relative and absolute)

This method considers the composition of the absolute feature (*count of distinct Bluetooth devices*) and the relative features (*mean signal strength, variance of the signal strength*). The idea is to maximize the classification performance metrics when combining absolute and relative features to a single feature vector.

#### 4.6.2 Group Formation Scanning Feature Vector

‘Group formation scanning’ (or ‘group scanning’) is an extension of individual scanning relying on multiple continuously nearby persons are roaming through the area. ‘Group scanning’ is defined as scanning with three mobile scanners in continuous proximity of 5 m displacement (triangle formation). The instantaneous



scan area is  $440m^2$ . The scanning range is extended as a side effect. The overlapping scanning range compensates ‘dead zones’ where some Bluetooth devices would otherwise be fully shielded at high crowd densities and not detected by an individual scanner.

### **Average count of crowd devices in the proximity (group absolute features)**

Groupwise scanning is defined as multiple people roaming in a triangle formation with  $5m$  displacement through the same space. The first group feature (*average number of Bluetooth devices*) is computed by collecting the *individual count of Bluetooth devices* of each participating scanning device and averaging the values. This feature is intended to compensate variations between scanners.

### **Set size of crowd devices in the proximity (group absolute features)**

The second group feature (*set size of Bluetooth devices*) is the compound set of actual surrounding devices to the group of people ( $count_{group}(i) = |\cup_{s \in i} \forall g|$  where  $s$  is a time of occurrence and  $i$  is the current interval and  $g$  is the group of scanners) in this time window independent to real duration of occurrence of a device. As opposed to a sum, this feature is independent of the number of collaborating devices.

### **Count variations between scanners (group relative count features)**

The third group feature (*variance in the number of devices*) is defined by the variance of the *individual count of Bluetooth devices* values across all people ( $variation_{count}(i) = var(count(i) \forall g)$ ).

## Signal strength variations (group relative RSSI feature)

The fourth group feature (*variance of all signal strengths*) is defined by the variance of the signal strengths aggregated from all participating scanning devices during a given scan interval ( $variance(i) = \frac{\sum_{x \in i} signal_x^2}{number_i} - average(i)^2 \forall g$ ). Potential multiple occurrences of the same Bluetooth device found by different sensing devices are not removed from the feature computation.

While the first and second group features represent absolute values, the third and fourth group features represent differences in the values measured by different, spatial distributed devices. Thus, they are more related to the properties of the crowd than to the absolute number of discoverable devices in the crowd (although they are not fully independent of the number of devices). The effects involved are complex and driven by a number of factors. For one, the spatial variance is likely to be reduced as the crowd density increases since each scan is likely to be based on a larger (= more representative) sample of people. On the other hand, with increased crowd density occlusions, reflections and other propagation effects are likely to play a bigger role. These depend on the specific configuration of people at scan locations (where the devices are worn, how they are occluded etc.) which means that variance will increase.

## Compound features (group relative and absolute)

This method considers the composition of the absolute feature and the relative features. The set of features includes the previous single scanner features (*count of distinct Bluetooth devices*), *mean signal strength*, *variance of the signal strength* from each of the individual scanners and extends the feature set with the group features (*average number of Bluetooth devices*), *set size of Bluetooth devices*, *variance in the number of devices*, *variance of all signal strengths*. The idea is to maximize the classification performance metrics when combining absolute and relative features to a single feature vector.

### 4.6.3 Machine Learning

We apply machine learning where the predicted outcome is the class  $Y$  to which the input data records belongs to. For classifier training we applied 10-fold stratified cross-validation. The set of data records is randomly partitioned into ten equal sized subsets while maintaining the same proportion of the target variables in each subset. A single subset is maintained as the validation data for testing the classifier model and two subsets are used for training the classifier. Cross validation is performed three times, while each data record is used exactly once for validation. All data records are used for training and validation during the three folds. The results are then averaged to produce a single estimation. As the machine learning classifier we used the decision tree learning algorithm. The decision tree learning algorithm is defined as the construction of a decision tree from training data records  $(x, Y)$ . The decision tree consists of branches representing conjunctions of 1 to  $k$  features leading to the leaves representing target variables  $Y$ . Each internal (non-leaf) node denotes a test on an variable  $x$ , each branch represents the outcome of a test, and each leaf (or terminal) node holds a target variable  $y_m$ . The topmost node in a tree is the root node. When predicting an unseen data record the tree is traversed from the root node to a leaf node representing the target variable  $y_m$ . The decision tree classification algorithm is configured as follows: the criterion measuring the split quality is set to the Gini impurity. The minimum number of samples required to split an internal node is set to 2. The minimum number of samples required to be at a leaf node is set to 1. A node will split if its impurity is above the threshold  $1e - 07$  (threshold for early stopping in tree growth).

We used performance metrics to evaluate the classifier. The classification ‘score’ is defined as the proportion of feature vectors  $\bar{x}$  exactly matching the corresponding label  $Y$ . The ‘precision’ performance metric is defined as  $\frac{\text{true positive}}{\text{true positive} + \text{false positive}}$  and can be described as the proportion of correctly predicted  $\bar{x}$  as target variable  $y_m$  to all  $\bar{x}$  predicted as  $y_m$ . The ‘recall’ performance metric is defined as  $\frac{\text{true positive}}{\text{true positive} + \text{false negative}}$  and can be described as the proportion of correctly predicted target variable  $y_m$  to the total number of correctly predicted target variables. The ‘f1-score’ performance metric is defined as  $2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}$  and can be interpreted as a weighted average of the precision and recall. The ‘confusion matrix’ performance

metric is defined as the matrix  $C$  such that  $C_{m,n}$  is equal to the proportion of data records known to belong to target variable  $y_n$  but predicted to be as target variable  $y_m$ .

#### 4.6.4 Experimental Environment and Data Collection Process

The following evaluation is based on the dataset collected during the Munich Oktoberfest event over a period of three days. The event was attended by hundreds of thousands of visitors from all over the world each day. The event's main pedestrian zone is approximately 500 m long and 20 m wide. The pedestrian zone is divided sparsely by merchandise stands in the center and bounded by food stands and tents with side street crossings where other visitors entered and left the area. Volunteers of the experiment were told to move continuously in a triangle group formation with a distance 5m to each other as accurate as this was possible in large crowd densities. Volunteers walked back and forth the main 500 meters long pedestrian event zone. After finishing one walk a short period of being stationary existed. For our experiment three participants were equipped with Android HTC Desire smartphones with enabled Bluetooth unit. The phones were placed in the pants front pockets. Collected raw data for each discovered Bluetooth device consisted of the following attributes:

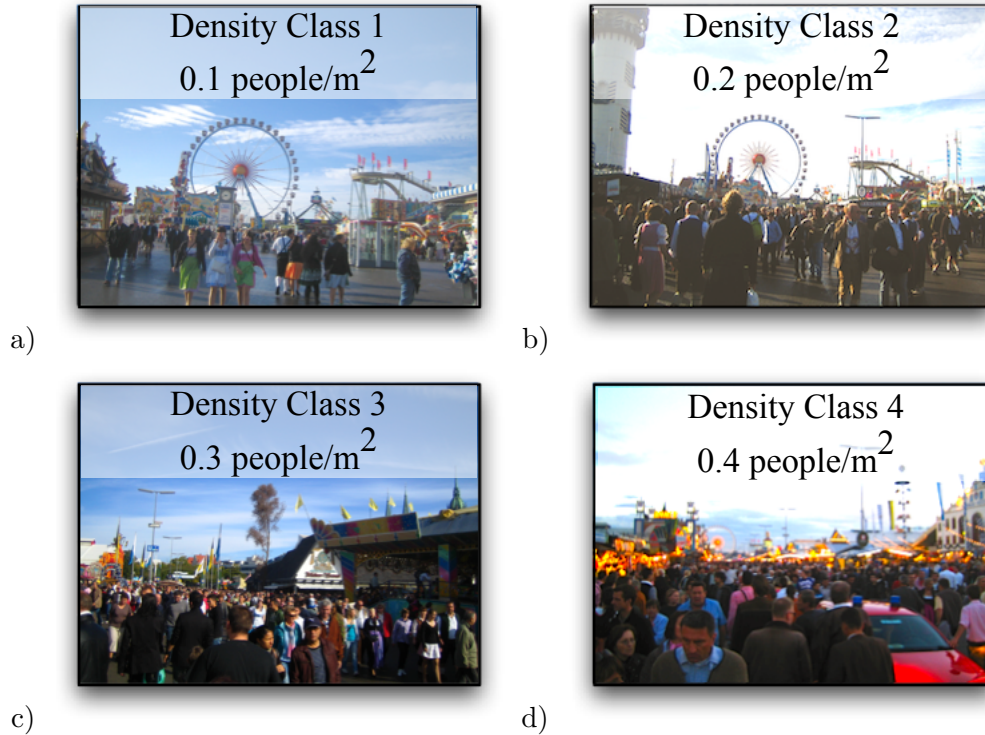
1. Timestamp
2. Bluetooth scan interval number
3. Bluetooth device name
4. Unique Bluetooth device ID
5. Bluetooth signal strength in dBm

The data was written to the SD card of each phone and evaluated later on. It is important to mention that the discovery process of Bluetooth devices is not an infinitely small snapshot in time, but in our case it was a 60 seconds time window (so called Bluetooth scan interval). During this time the underlying system reports Bluetooth devices which were not found before in the given scan interval. Therefore the above features are based on a time interval of about 60 seconds.

For crowd density ground truth one group member took digital camera pictures at an interval of 100 m along the zone. All pictures had an embedded timestamp and were made with the same focal length of 5.8 (equivalent to 35 mm) and consistent angle of vision while holding the camera above the head in forward direction. The area regarded for manual ground truth count of the people per picture was defined by the maximum distance where the head width falls below 10% of the picture width. With this technique some heads might have been covered by other heads, but compared to birds-eye-perspective-pictures this approach was feasible in this situation. All pictures were annotated manually according to the defined boundaries and averaged over pictures taken per walk on the 500 m zone resulting in a crowd density label for each segment of the experiment. The crowd condition is measured in terms of crowd density which varied between 0.1 and 0.4  $m^{-2}$  with a resolution of 0.1  $m^{-2}$  (see Figure 4.4). Measurement evaluation yielded that 5–7% of the people have Bluetooth enabled on their mobile phones during the Munich Oktoberfest data set experiment.

#### 4.6.5 Data Distribution

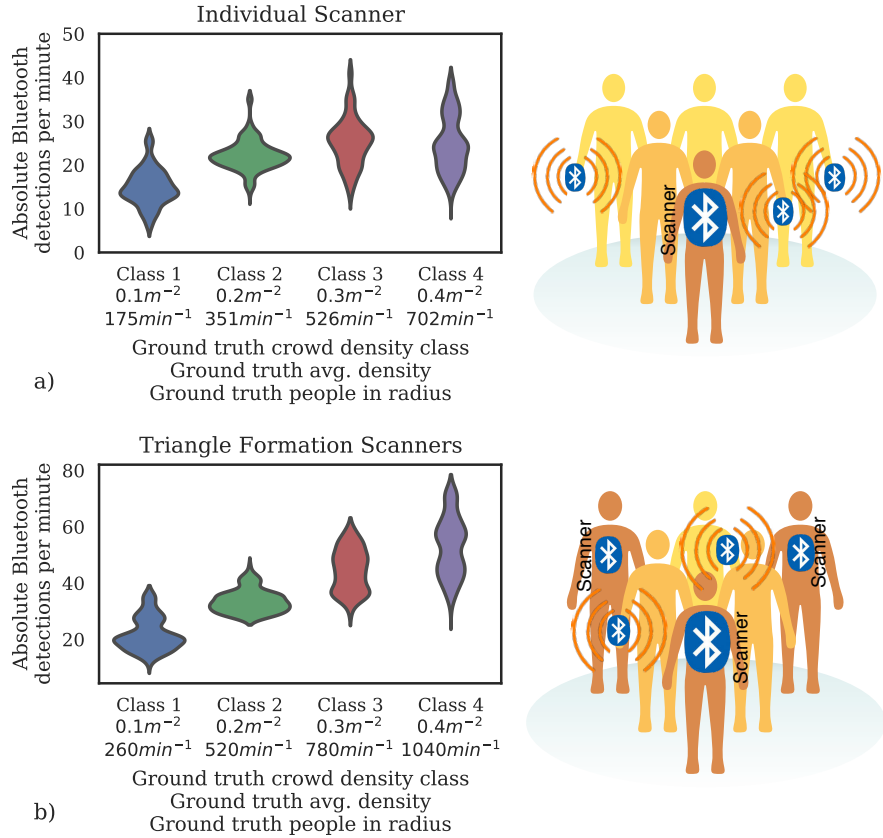
The general considerations in section 4.5 included the issue of the question of signal attenuation. Due to signal attenuation it is expected that the number of detected devices does not linearly increase. To validate this expectation, we analyzed the absolute count of discovered devices by an individual scanner and by a group of scanners within the Bluetooth proximity. In the violin plot in Figure 4.5a we see the ‘individual scanner’ distribution of the absolute count of distinct Bluetooth devices per minute during constant motion. The ground truth is shown as the crowd density class 1 to class 4, the crowd density [ $m^{-2}$ ], and the number of people in the 10 m range radius (based on mixed crowd flow in both directions). Four different crowd densities have been detected during the experimental evaluation by ground truth: 0.1  $m^{-2}$  (class 1), 0.2  $m^{-2}$  (class 2), 0.3  $m^{-2}$  (class 3), and 0.4  $m^{-2}$  (class 4). We observe that the increasing ground truth crowd densities does not linearly correspond to the distributions of measurements for each class. The properties of the distribution of class 3 and class 4 are similar. The peak of the class 4 distribution even falls below the class 3 distribution, although more people being



**Figure 4.4.** Ground truth reference images representing average crowd densities of  $0.1m^{-2}$ ,  $0.2m^{-2}$ ,  $0.3m^{-2}$ , and  $0.4m^{-2}$ .

in the proximity. This affirms our expectations and demonstrates the issue for crowd density estimation with the naive approach of a single mobile sensor and pure counting of devices in the proximity.

In violin plot in Figure 4.5b we see the ‘group of scanners’ distribution of the set size of Bluetooth devices per minute during constant motion. The ground truth is shown as the crowd density class 1 to class 4, the crowd density [ $m^{-2}$ ], and the number of people in the  $440m^2$  triangle formation scan area with partial scan range overlapping between scanners. We observe that increasing ground truth crowd densities correspond to increasing peak values of the distributions. The distributions still overlap partially between the crowd density classes especially between class 3 and class 4. However, the distinction between the distributions is superior to the distributions in Figure 4.5a.



**Figure 4.5.** Violin-plot data distribution of the feature absolute number of unique Bluetooth device detections against ground truth. Ground truth is shown as the crowd density class, the crowd density and the number of people in 10 meters radius (based on mixed crowd flow in both directions). The distribution plots show a) individual mobile scanning, and b) group scanning which is defined in this case as scanning with 3 mobile scanners in a triangle formation with 5 m displacement and overlapping scanning range.

### 4.6.6 Experimental Validation Results

We present the results of the machine learning evaluation for seven feature vectors listed in Table 4.4. The feature vector were constructed without further windowing the data. The estimation is performed in real-time with the delay of one scanner interval duration of 60 seconds. When applying the machine learning to the one dimensional feature vector of the *individual absolute feature* we achieve a classi-

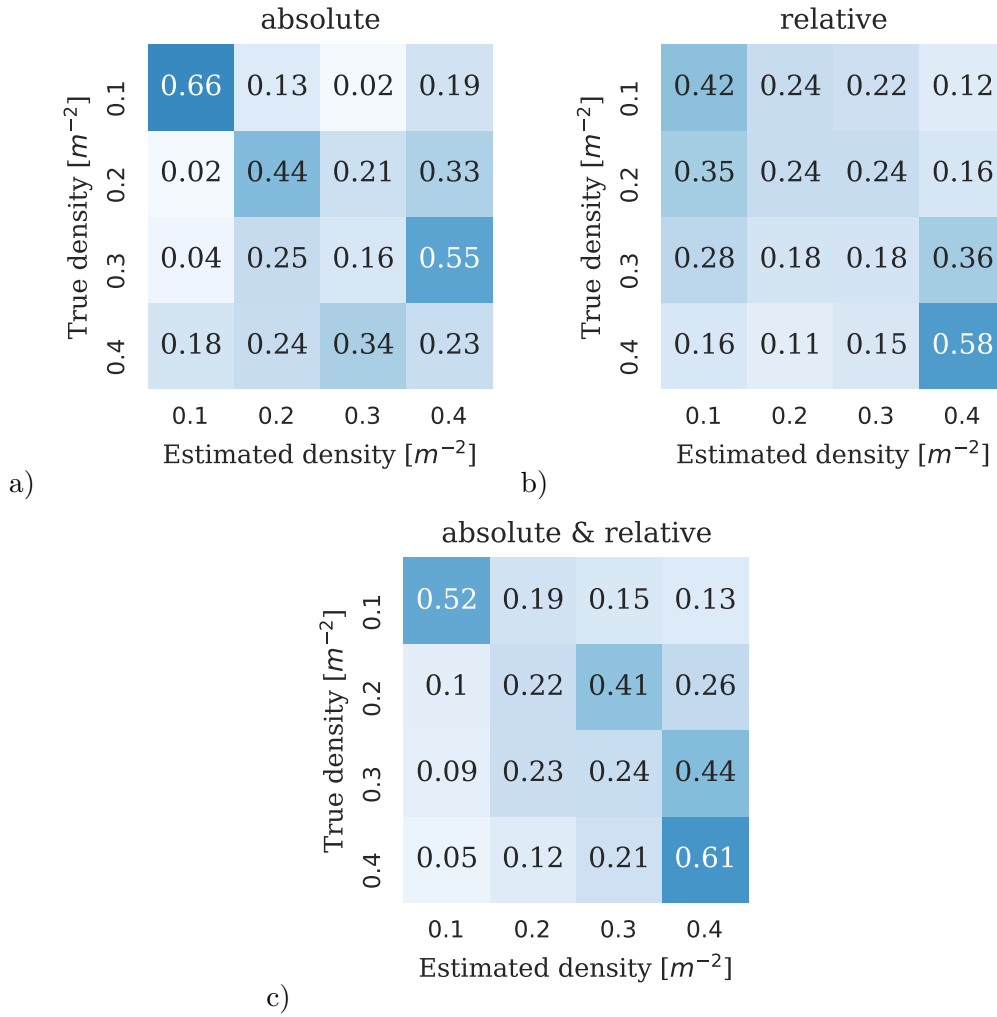
**Table 4.4.** Classification metric performance results of crowd density estimation with mobile individual scanners and with a single scanning group

Feature Vector	Classification results			
	Score	Precision	Recall	F1-Score
individual absolute	0.34	0.35	0.34	0.33
individual relative RSSI	0.39	0.40	0.39	0.39
individual absolute & relative	0.44	0.45	0.44	0.43
group absolute	0.65	0.66	0.65	0.64
group relative RSSI	0.55	0.55	0.55	0.54
group relative COUNT	0.30	0.30	0.30	0.29
group absolute & relative	0.66	0.67	0.66	0.65

fication score of 0.34 (see further performance metrics in Table 4.4), while the mis-classifications occur across all classes in the confusion matrix (Figure 4.6a). Class 0.4 is more often mis-classified as class 0.3 or even class 0.2. Class 0.3 is classified as class 0.4 by 55%. When applying the machine learning to the two dimensional feature vector of the *individual relative features* we achieve a classification score of 0.39, while the mis-classifications occur across all classes in the confusion matrix (Figure 4.6b). Classes 0.1 and 0.4 are estimated best, while the embraced classes 0.2 and 0.3 are estimated with the score of a uniformly guessing or even below. When applying the machine learning to the three dimensional feature vector of the individual compound feature vector of the *compound individual absolute and relative features* we achieve a classification score of 0.44, while the mis-classifications occur across all classes in the confusion matrix (Figure 4.6c). Compared to the previous feature vectors this equals to an increase in the classification score of 0.10 or 0.05.

The feature vector *group absolute features* consists of all group absolute and the inherited individual absolute features. By combining similar features of the same category we improve the clarity of the evaluation and improve the estimation performance with multi-dimensional feature vectors. When evaluating the machine





**Figure 4.6.** Confusion matrices representing the class-wise machine learning performance metric classification results of the individual scanning approach. Confusion matrices of classification results with a) individual scanners with absolute feature, b) individual scanners with relative features, and c) individual scanners with absolute feature and relative features.

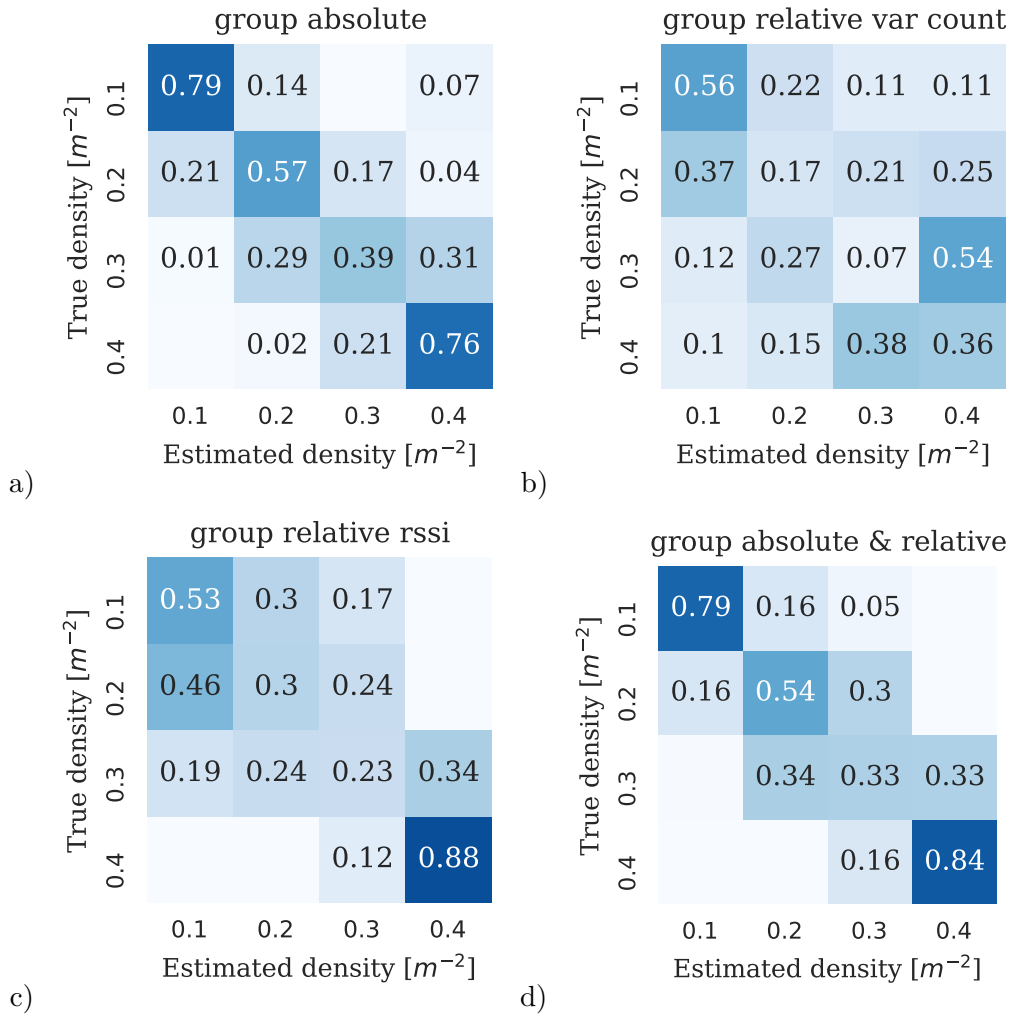
learning results on the 5-dimensional feature vector of the *group absolute features* we achieve a classification score of 0.65, while the mis-classifications mostly occur along the diagonal at neighbored classes in the confusion matrix (Figure 4.7a). Especially class 0.3 is often confused with class 0.2 or 0.4. When a crowd density

application allows a class divergence of  $\pm 1$ , the classification score is between 0.93 and 0.98 per class for the now estimated class range.

The feature vector *group relative count feature* consists of a single feature. When applying the machine learning to the 1-dimensional feature vector of the *group relative count feature* we achieve a classification score of 0.30, while the mis-classifications occur across the whole confusion matrix (Figure 4.7b). The classes 0.1 and 0.4 are estimated with 0.56 and 0.36, while the intermediary classes 0.2 and 0.3 are estimated with a significant lower score of 0.17 and 0.07.

The feature vector *group relative RSSI features* consists of the group relative feature and the inherited *individual relative RSSI features*. When evaluating the machine learning results on the 7-dimensional feature vector of the *group relative RSSI features* we achieve a classification score of 0.55, while the mis-classifications mostly occur along the diagonal at neighbored classes in the confusion matrix (Figure 4.7c). The classes 0.1 and 0.4 are estimated with 0.53 and 0.88, while the intermediary classes 0.2 and 0.3 are estimated with a significant lower score of 0.30 and 0.23. When a crowd density application allows a class divergence of  $\pm 1$ , the classification score -for the now estimated class range- is between 0.27 and 1.0 per class.

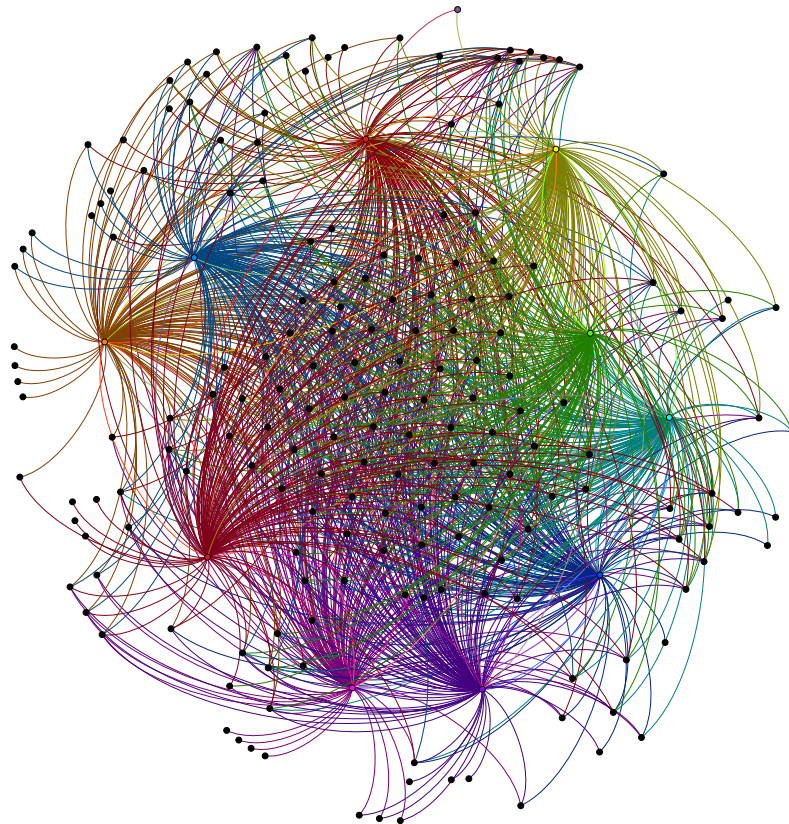
The feature vector *group compound relative and absolute features* consists of the individual absolute feature, the individual relative features, the group absolute features, the group relative RSSI features and the group relative count feature. When evaluating the machine learning results on the 13-dimensional feature vector of the *group compound relative and absolute features* we achieve a classification score of 0.66, while the mis-classifications occur just along the diagonal at neighbored classes in the confusion matrix (Figure 4.7d). When a crowd density application allows a class divergence of  $\pm 1$ , the classification score -for the now estimated class range- is between 0.95 and 1.0 per class.



**Figure 4.7.** Confusion matrices representing the machine learning performance metric of mis-classifications of the group scanning approach. Confusion matrices of classification results a) group scanning and absolute feature, b) group scanning and relative features, and c) group scanning and absolute feature and relative features.

## 4.7 Collaborative Scanning - Public Viewing Experiment

We introduce new collaborative concepts of multiple teams walking intermittently nearby and scanning each other in addition to the previous method. The main contributions beyond the methods presented in section 4.6 include processing information about the link structure between actively scanning Bluetooth devices, ratio of discovered devices in the current scan window to previous scan windows, teamwise diversity of discovered devices, and the number of semi-continuous device visibility periods. Additionally, it takes device visibility durations into account.



**Figure 4.8.** Bluetooth link structure graph showing all 10 actively scanning smartphones (colored nodes) and discovered devices (black nodes) during the experiment.

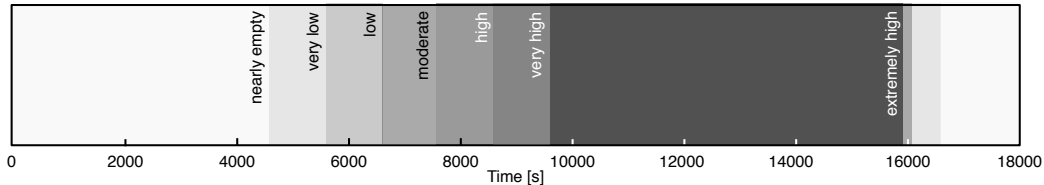
### 4.7.1 Experimental Environment and Experimental Setup

To evaluate the methods described in the following sub-section we set up three experiments on three days during the European soccer championship 2012 at the official public viewing event at the town-center (marketplace called ‘Stiftsplatz’) of Kaiserslautern (Germany). The evaluated experiment area has a dimension of 48.5 to 48.5 meters allowing up to 5200 people to enter the fenced area.

Each experiment had a duration of about 4 hours consisting of 2 hours before the soccer championship kick-off began, 45 minutes during the first half of the soccer match, 15 minutes during the half-time break, 45 minutes during the second half of the soccer match, and 20 minutes after the game.

We started our experiment early before spectators began entering the event area. During two hours the area was then filled up to a level where no more people were allowed to enter the area by the event organization for safety reasons.

We gathered sensor data of different crowd densities including levels *nearly empty* ( $0.01 - 0.05\text{people}/\text{m}^2$ ), *very low* ( $0.05 - 0.2\text{people}/\text{m}^2$ ), *low* ( $0.2 - 0.3\text{people}/\text{m}^2$ ), *moderate* ( $0.3 - 0.4\text{people}/\text{m}^2$ ), *high* ( $0.4 - 1.0\text{people}/\text{m}^2$ ), *very high* ( $1.0 - 2.0\text{people}/\text{m}^2$ ), *extremely high* ( $2.0 + \text{people}/\text{m}^2$ ). See figure 4.9 for the complete course of the crowd density levels during the experiment and figure 4.11 with excerpts from the ground truth video for each crowd density class.

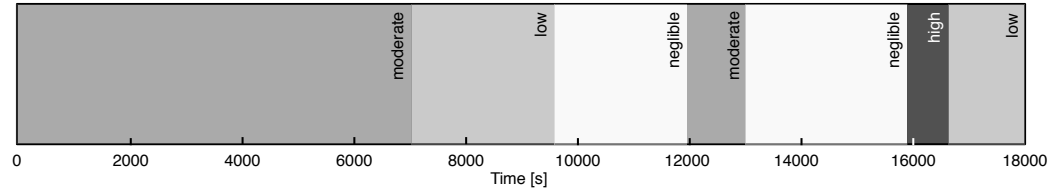


**Figure 4.9.** Crowd density ground truth information during the course of the experiment.

The crowd flow was moderate during the filling phase of the event since attendees went slowly from the event entry towards the big screen in the opposite corner. There was no crowd flow during the first and second half of the soccer match. During the break the crowd flow range was between moderate and high. After the end of the soccer match the crowd flow was very high since the attendees wanted to leave the event area through multiple exits as fast as possible because the German

#### 4.7 Collaborative Scanning - Public Viewing Experiment

soccer team had lost the match. See figure 4.10 for the complete course of the crowd flow. Our presented crowd density measurement technique is robust to differing crowd flow levels as they are not considered in the labeling procedure of the feature vectors and are not correlated to the crowd density levels.



**Figure 4.10.** Crowd flow ground truth information during the course of the experiment.

For each experiment we recruited 10 students. We divided the students into 5 teams with 2 students each. Team members always stayed in close contact (up to 1 meter distance) to each other. Teams were instructed to be either stationary (2 teams, 4 students) or dynamic (3 teams, 6 students).

*Stationary* is defined as continuously standing on the spot. We placed stationary teams near the entrance of the area. One team near the left side and the other team on the right side of the entry.

*Dynamic* is defined as continuously walking around on the event area. Teams were told to walk on a curved path covering 3 sides of the event area and mostly covering the edge regions (see below for exceptions) of the crowd since those regions were common to walk on because of nearby food and beverage stands.

The idea behind the stationary and dynamic scripted setup is to represent a natural behavior of people during such events. Some people are standing still watching the performance. Other people are walking around to food/beverage stands, meet friends, change to a better viewing spot etc.

Multiple dynamic teams are allowed to walk asynchronously. The walking speed of dynamic teams is not scripted, allowing to choose the personal optimum walking speed (we can determine the walking speed by evaluating our GPS log information). We allowed teams to move to a self-determined place in the middle of the crowd during the first and second half of the soccer match excluding the break.

In a real-life scenario people do not have to be categorized to behave stationary or dynamically continuously. Smartphone sensor information allows to dynamically

detect the type of behavior. Because of this random natural behavior we do not manually apply any information to our algorithms about stationary or dynamic behavior.

Each student was equipped with one Android smartphone which is placed in the trouser's pocket.

We deployed Android smartphones of different types including HTC Desire, Google Nexus and Samsung Nexus S each based on the most recent version of the Android operating system. On all devices we were running our custom Android application called ContextRobot which records multiple sensor data streams onto the microSD card for later off-line analysis. Our Android applications continuously scanned for discoverable Bluetooth devices (a Bluetooth scan is defined as a time interval which emits a set of unique Bluetooth devices with the restriction of unrepeated occurrences of a unique device). A single log entry of a device discovery during a Bluetooth scan is associated with a timestamp, a serial Bluetooth scan interval number, personal Bluetooth device name, unique Bluetooth identifier (Bluetooth MAC address), and the Bluetooth signal strength as a RSSI (received signal strength indication) value.

An exact temporal begin and end of a Bluetooth scanning interval cannot be specified during the recording of Bluetooth sensor data since the operating systems restricts to certain length of scan intervals depending on internal thresholds. The average duration of a Bluetooth scan interval is about ten seconds (with little variations). Our application triggers a new Bluetooth scan when the previous scan has ended. Multiple collaborative Android devices recording Bluetooth scans intervals are synchronized in an off-line manner. At a given time window of a length of 20 seconds we determine one scan interval which fits this window entirely. In addition to Bluetooth sensor information we record location information by the GPS sensor at a frequency of 1 Hz. Our Android application continuously records timestamp, latitude, longitude and accuracy onto the microSD cards. Location information is required for some feature computations which rely on distances between multiple students and their walking speed.

For obtaining ground truth data about the crowd density we set up a HD video camera on top of a neighbored hotel building with view of the whole event area. Figure 4.11 shows excerpts of the video footage for different crowd density classes.

The ground truth labels are based on the video footage which we labeled every 10-15 minutes with a crowd density class ranging from nearly empty to extremely high.

### 4.7.2 Feature Vector

We are calculating our features based on multiple partially distributed sensors because we want to achieve a statement of the crowd density of the whole event area as we assume all sensors together are covering a large portion of the area during movement in the area.

#### **Feature: Averaged sum of distinct devices discovered by all sensors in scan window**

This simple feature describes the current number of discovered distinct devices for every snapshot (a *Bluetooth scan window* is hereafter also referred to as a snapshot) of the experiment. For each snapshot each of the sensors delivers a set of unique devices identified by the unique Bluetooth MAC address.

Calculating the union of all discovered devices (by all sensors) divided by the number of sensors results in this feature. Bluetooth devices discovered by multiple sensors at the same time are not influencing this feature.

This feature relies directly on the level of distribution of the sensors over the event area. Since the Bluetooth range is very limited a sensor distribution over a larger area obviously leads to a larger number of distinct devices and the other way around.

The downside of this feature is its direct relation to the ratio of discoverable Bluetooth devices to the number of people to be detected.

See figure 4.12 on page 123 for a visualization of the feature during one experiment.



**Feature: Ratio bi-directional link structure of sensors to average pairwise distance of sensors multiplied with average sensor speed**

This composite feature characterizes the context of the snapshot of the collaborative sensor data more explicit.

The feature takes the bi-directional Bluetooth link structure between the sensors (actively scanning *Bluetooth devices* are hereafter also referred to as sensors) into account. A directional link (hereafter also referred to as sensor discovery) between a pair of sensor ‘a’ and ‘b’ is defined as established when sensor ‘a’ discovers sensor ‘b’ in the current snapshot. Another directional link is defined as established when sensor ‘b’ discovers sensor ‘a’. This implies a pair of sensors might both link to each other or one sensor links to the other or none of both sensors links to the other. All combinations of pairs between sensors are monitored. Maximum established links between ten sensors would be 90 links, the minimum number of established links would be zero. The bi-directional link structure is defined as the sum of all links between all sensors.

The average pairwise distance calculation is based on the GPS sensor data information. A snapshot contains multiple GPS locations per sensor (GPS location is sampled at 1 Hz). Only locations with a GPS accuracy better than 15m are taken into account. Based on the filtered locations we calculate the average as most promising location of the sensor. For each pair of sensors we calculate the distance between them. There are  $n!/2/(n-2)!$  distances calculated per snapshot where  $n$  is the number of sensors. The distance between all sensor pairs is then averaged. The average speed is calculated for every snapshot and each individual sensor based on averaged GPS information per snapshot each with an accuracy of better than 15m. Afterwards the average speed is calculated for all sensors. Finally, the feature is calculated by the number of bi-directional links divided by the average pairwise distance of sensors multiplied with the average sensor speed. It is important to mention that this feature is completely independent of external (others than the used Bluetooth sensors) discoverable Bluetooth devices. It uses the relationship between the number of links to the distance between sensors, based

on the assumption that a more dense crowd shields the sensor links heavier than in a low dense crowd with the same underlying distance.

See figure 4.13 on page 124 for a visualization of the feature during one experiment.

### **Feature: Ratio of discovered devices in current snapshot to discovered devices in last x minutes**

This feature characterizes the crowd movement during the snapshot of the collaborative sensor data more explicit.

Newly detected Bluetooth devices in a snapshot are defined as a set of all unique devices discovered during the snapshot by all sensors. Calculating the union of unique discovered devices by all sensors in a snapshot leads to the collaborative measurement. The second part of the ratio is the size of the set of unique Bluetooth devices discovered during previous 15 snapshots (depending on the size of the snapshot, this signifies a monitoring of the previous 1 to 10 minutes). Finally, the feature is calculated by the size of the collaborative set of discovered devices at the snapshot divided by the size of the collaborative set of Bluetooth devices discovered before. This implies that the value is smaller in a less moving crowd than in a more likely moving crowd. This is caused by the fact that the number of different devices seen during x snapshots is smaller if there is less movement (less devices are rushing by) than for strong crowd movement (high crowd flow).

See figure 4.14 on page 124 for a visualization of the feature during one experiment.

### **Feature: Average *teamwise* diversity of discovered devices per scan window (ratio not concurrent devices to concurrent devices)**

We define a team by two persons staying in close adjacency while each person is carrying a sensor. A team can either move dynamically or be stationary, but continuously stays together.

This feature takes into account the *teamwise* diversity of discovered Bluetooth devices for each snapshot. In this context we define diversity as the ratio between

the number of Bluetooth devices not concurrently discovered and concurrently discovered devices. Not concurrently discovered devices are defined in set theory as the symmetric difference. Either sensor ‘a’ or sensor ‘b’ but not both sensors discover the same device in a snapshot. Concurrent devices appear both in the current snapshot of sensor ‘a’ and sensor ‘b’. The ratio is averaged for all teams in each snapshot for a collaborative measurement.

This feature calculates the diversity of discovered devices between two sensors which are close to each other. This gives us a value depending on the crowd between and around the team as well as the unambiguousness of the two sensor measures.

See figure 4.15 on page 125 for a visualization of the feature during one experiment.

### **Feature: Average number of semi-continuous unique device visibility periods (finite state machine approach)**

We define a semi-continuous device visibility period per sensor as the number of consecutive snapshots, whereas in each snapshot a unique device is discovered with the exception of very short vanishings during the period. A *short vanishing* is defined as a single snapshot without a discovery among other snapshots including the presence of a specific device. Multiple short vanishings may appear during a semi-continuous device visibility period. The period ends when a device vanishes at least for two consecutive snapshots. The same unique device then may reappear again or vanish for a longer time or forever.

The data is further processed by calculating the sum of present semi-continuous unique device visibility periods during a snapshot. By definition, the sum of unique devices might include a device which is not seen in the current snapshot. We implemented the calculation of this feature by a finite state machine for each unique device (d) and for each sensor (s). Resulting in  $d * s$  finite state machines. Finally, the collaborative overall average value is calculated per snapshot over all sensors. The feature value can be high for a small number of discoverable devices which are in range for a longer time. The value can be low for a high number of discoverable devices which are in range for a shorter time.

See figure 4.16 on page 125 for a visualization of the feature during one experiment.

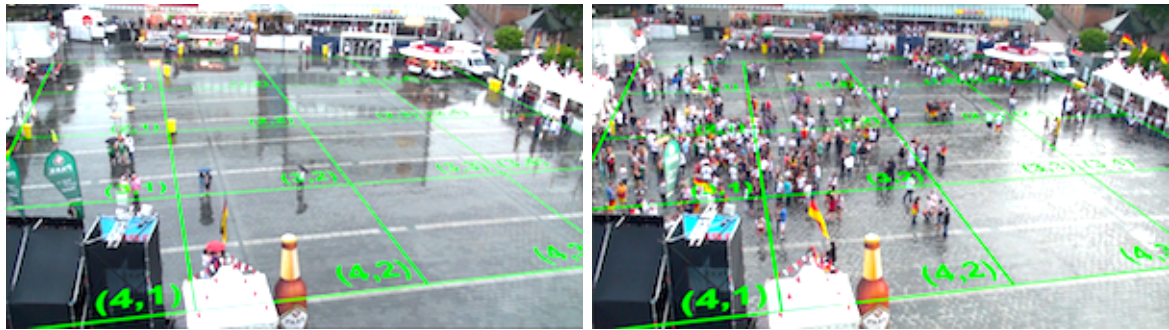
**Feature: Average durations of semi-continuous unique device visibility periods (finite state machine approach)**

This feature is based on semi-continuous device visibility similar to feature 4.7.2 but calculates the duration. Therefore, the pre-processing is similar to feature 4.7.2.

The duration of a semi-continuous visibility of a unique device is defined as the number of sequential snapshots where a specific device is seen. This duration factors into all snapshots that the semi-continuous visibility is covering. Averaging the duration for one sensor of all device visibility durations at one snapshot is the value per sensor. Averaging this value of all sensors per snapshot results in the value of this feature.

See figure 4.17 on page 126 for a visualization of the feature during one experiment.

4 Crowd Condition Estimation with Mobile Scanners Opportunistically Scanning Crowd Devices



(a) nearly empty  
0.01-0.05  $people/m^2$

(b) very low  
0.05-0.2  $people/m^2$



(c) low  
0.2-0.3  $people/m^2$



(d) moderate  
0.3-0.4  $people/m^2$



(e) high  
0.4-1.0  $people/m^2$



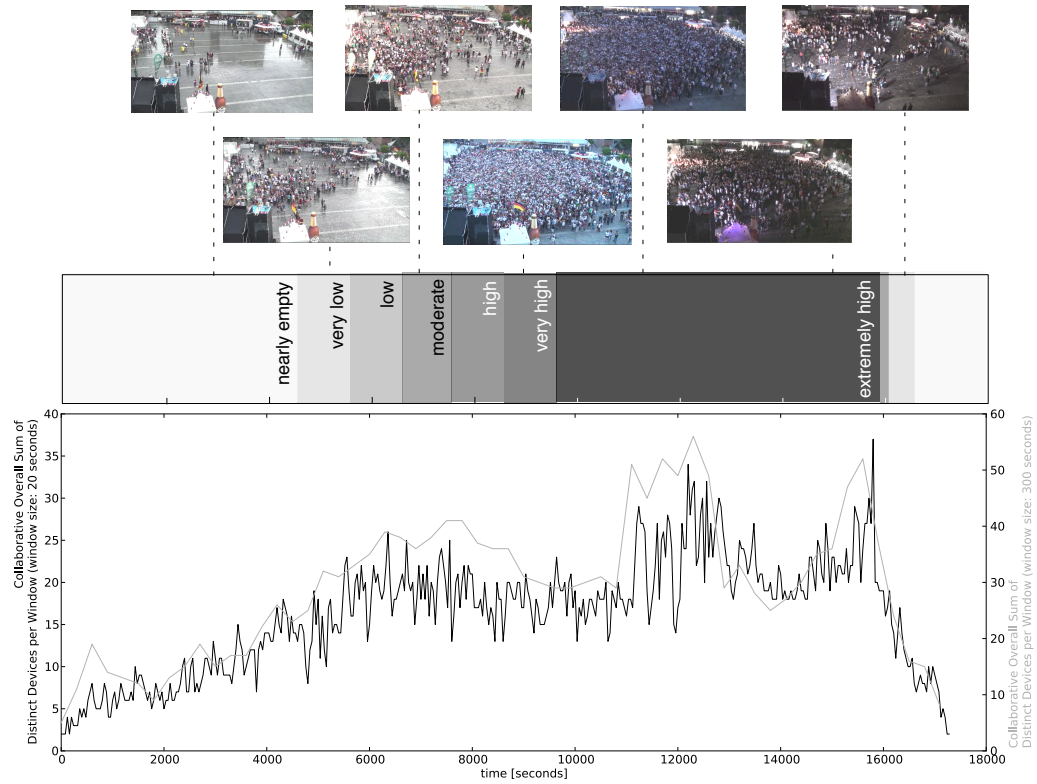
(f) very high  
1.0-2.0  $people/m^2$



(g) extremely high  
2.0++  $people/m^2$

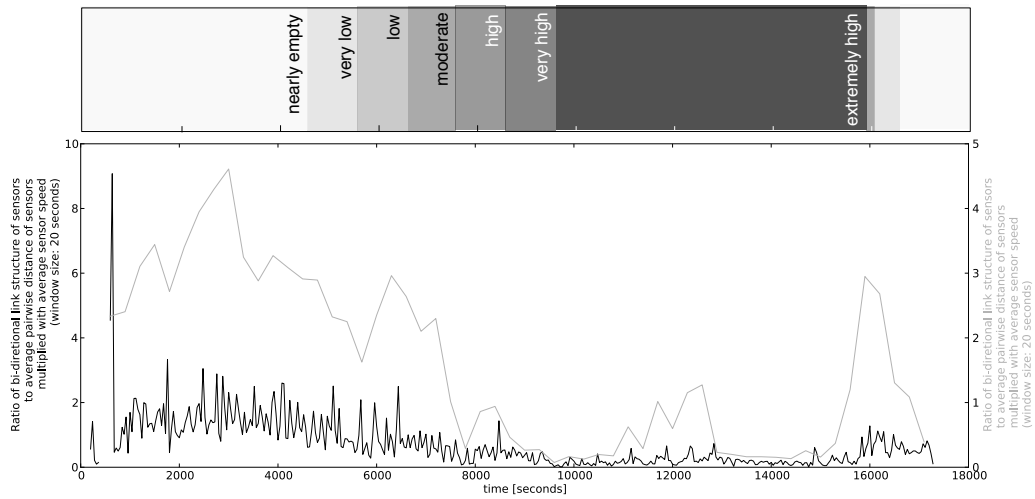
Figure 4.11. Crowd density classes ranging from *nearly empty* to *extremely high*. Excerpts of the HD ground truth video.

#### 4.7 Collaborative Scanning - Public Viewing Experiment

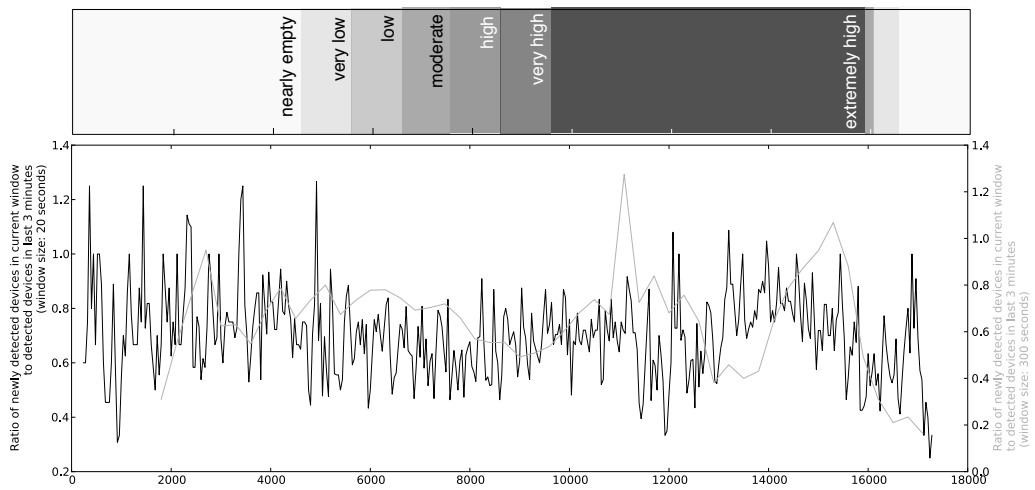


**Figure 4.12.** Feature: 'Size of device set of all distinct discovered devices by all sensors in time frame'. Overview of crowd density levels shown by different background grey levels.

#### 4 Crowd Condition Estimation with Mobile Scanners Opportunistically Scanning Crowd Devices

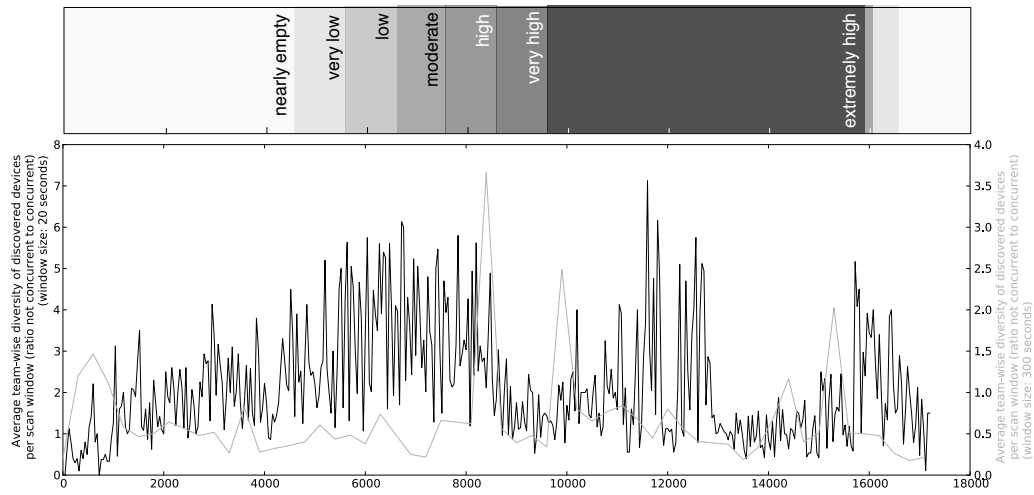


**Figure 4.13.** Feature: ‘Ratio bi-directional link structure of sensors to average pairwise distance of sensors multiplied with average sensor speed’

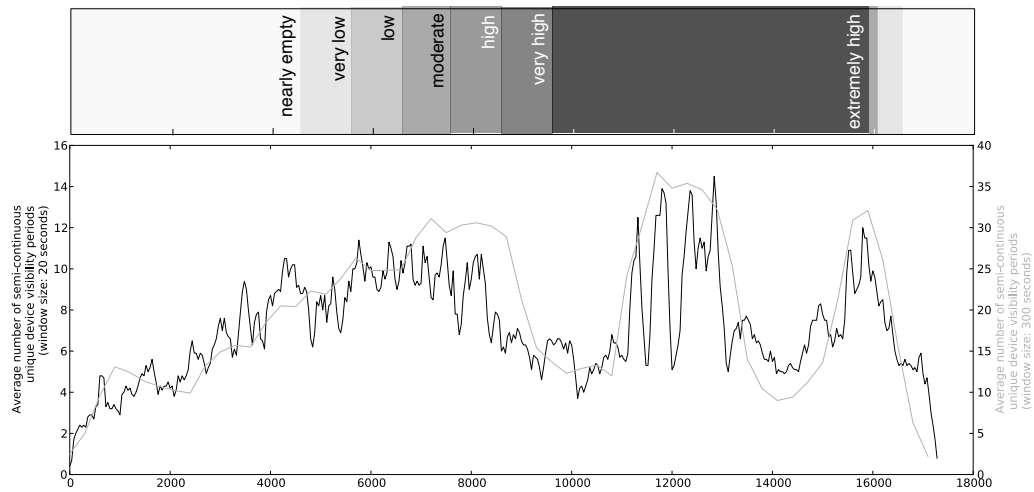


**Figure 4.14.** Feature visualization of ‘Ratio of discovered devices in current snapshot to discovered devices in last x minutes’.

#### 4.7 Collaborative Scanning - Public Viewing Experiment



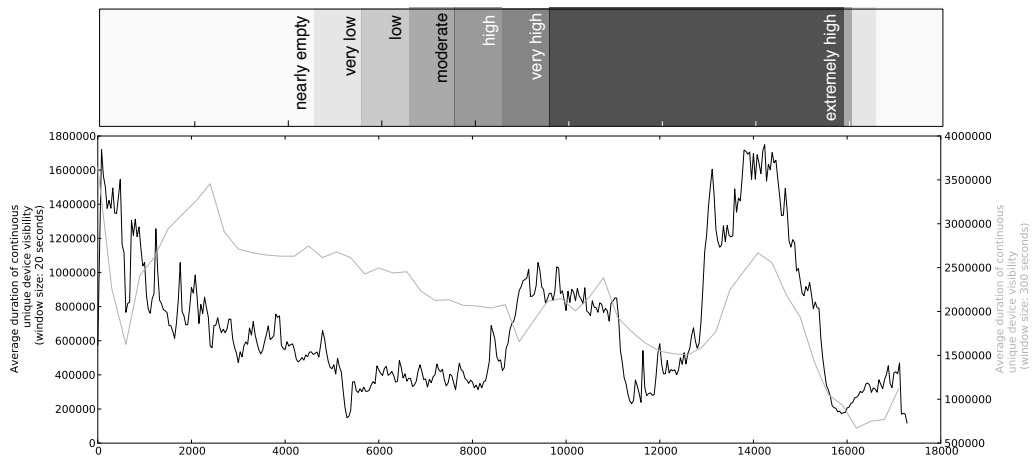
**Figure 4.15.** Feature visualization of ‘Average team-wise diversity of discovered devices per scan window (ratio *not* concurrent to concurrent)’.



**Figure 4.16.** Feature visualization of ‘Average number of semi-continuous unique device visibility periods’.



#### 4 Crowd Condition Estimation with Mobile Scanners Opportunistically Scanning Crowd Devices



**Figure 4.17.** Feature visualization of ‘Average durations of semi-continuous unique device visibility periods (finite state machine approach)’.

### 4.7.3 Experimental Validation

We introduced a new collaborative concepts of multiple teams walking intermittently nearby and scanning each other in addition to the previous method. The main contributions beyond the related work not relying just on the number of devices seen by a Bluetooth scan, but also take information about the link structure between actively scanning Bluetooth devices, ratio of discovered devices in the current scan window to previous scan windows, teamwise diversity of discovered devices, number of semi-continuous device visibility periods, and device visibility durations into account.

We validate the proposed collaborative features and compare them to the reference method of individual sensing. The different approaches of relying either on only absolute features (based on the count), relative features (based on relative and differential measurements between scanners), or on the combination of relative and absolute features are compared. ‘Collaborative absolute features’ are consisting of the two methods ‘averaged sum of distinct devices discovered by all sensors in scan window’ and ‘average number of semi-continuous unique device visibility periods (finite state machine approach)’. ‘Collaborative relative features’ are consisting of the four methods ‘ratio bi-directional link structure of sensors to average pairwise distance of sensors multiplied with average sensor speed’, ‘ratio of discovered devices in current snapshot to discovered devices in last 5 minutes’, ‘average teamwise diversity of discovered devices per scan window (ratio not concurrent devices to concurrent devices)’, and ‘average durations of semi-continuous unique device visibility periods (finite state machine approach)’. ‘Collaborative relative & absolute features’ are consisting of the six methods being a combination of the ‘collaborative absolute features’ and the ‘collaborative relative features’. The reference method is just building on individual scanner features as described in section 4.6.

As described earlier the approach with just relative features allows potential crowd density estimation applications without adapting the estimator to the saturation of mobile wireless devices. The crowd density estimation is evaluated with the machine learning method of classification with a real-time resolution of just 40 seconds. We selected the machine learning ensemble method ‘random decision tree forest classifier’ because it has been proven to handle complex data sets with a

multitude of features very well. We trained the ensemble random decision tree forest classifier with the common proven parameters. The classifier was cross-validated with 10-fold cross-validation. For cross-validation the stratified cross-validation was selected to achieve a balance in the classes used for training and estimating the classifier, due to the higher fraction of ‘extremely high’ crowd density class instances.

#### 4.7.4 Results

We achieve a classification score of 67% based on absolute collaborative features for estimating the correct crowd density class on seven discrete crowd density classes with stratified 10-fold cross-validation (see Table 4.5). This outperforms the reference method of absolute individual features by 27% (reference method score 40%). The precision is 0.71, the recall 0.69. We evaluate the mis-classifications of the classifier by visualizing the classifications results in a confusion matrix which presents the fraction of instances of classes being estimated as the true class or other classes. In the confusion matrix we observe mis-classifications between classes not directly adjacent to the true crowd density class. Classes ‘0.2 - 0.3’, ‘0.3 - 0.4’, ‘0.4 - 1.0’, and ‘2.0+’ are mis-classified in some instances to not similar classes. The different crowd density classes are classified with a classification score between 0.33 and 0.8. Figure 4.18 on page 130 shows the confusion matrix of the collaborative absolute features.

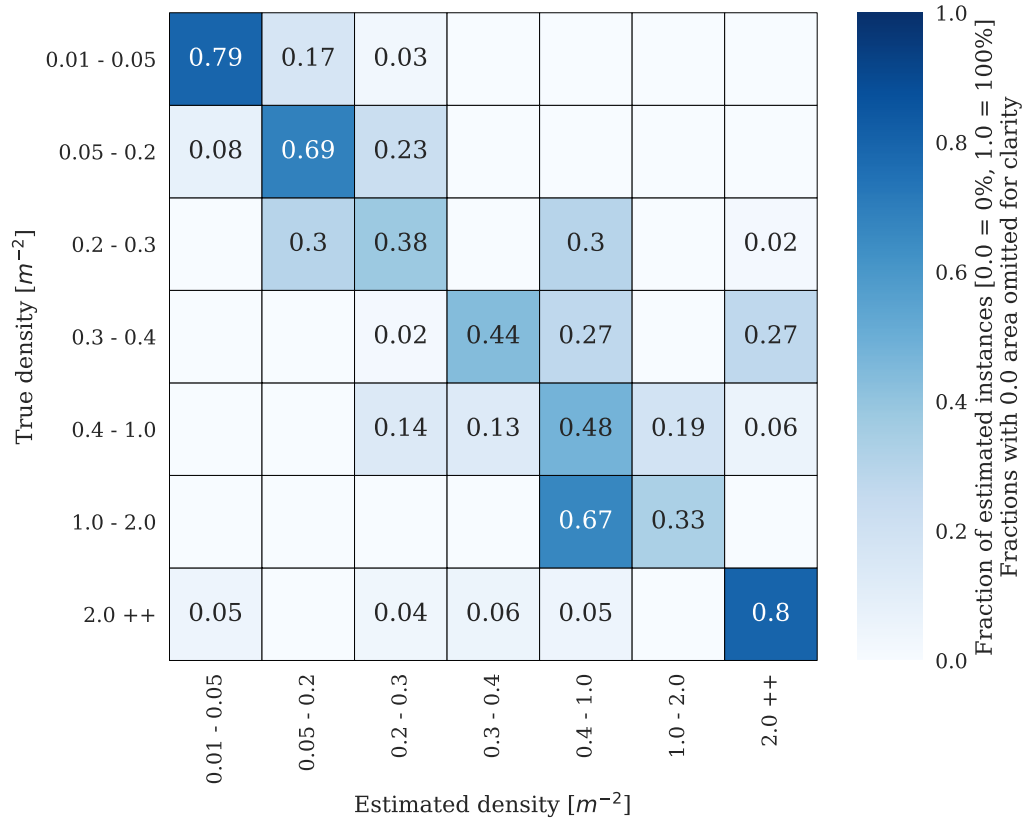
We achieve a classification score of 74% based on relative collaborative features for estimating the correct crowd density class on seven discrete crowd density classes with stratified 10-fold cross-validation (see Table 4.5). This outperforms the reference method of absolute individual features by 17% (reference method score 57%). In the confusion matrix we observe mis-classifications just between directly adjacent classes with few exceptions to the true crowd density class. Only classes ‘0.4 - 1.0’, and ‘2.0+’ are mis-classified just 3–4% of the instances to not similar classes. The different crowd density classes are classified with a classification score between 0.58 and 0.9. Figure 4.19 on page 131 shows the confusion matrix of the collaborative absolute features.

We achieve a classification score of 77% based on absolute and relative collaborative

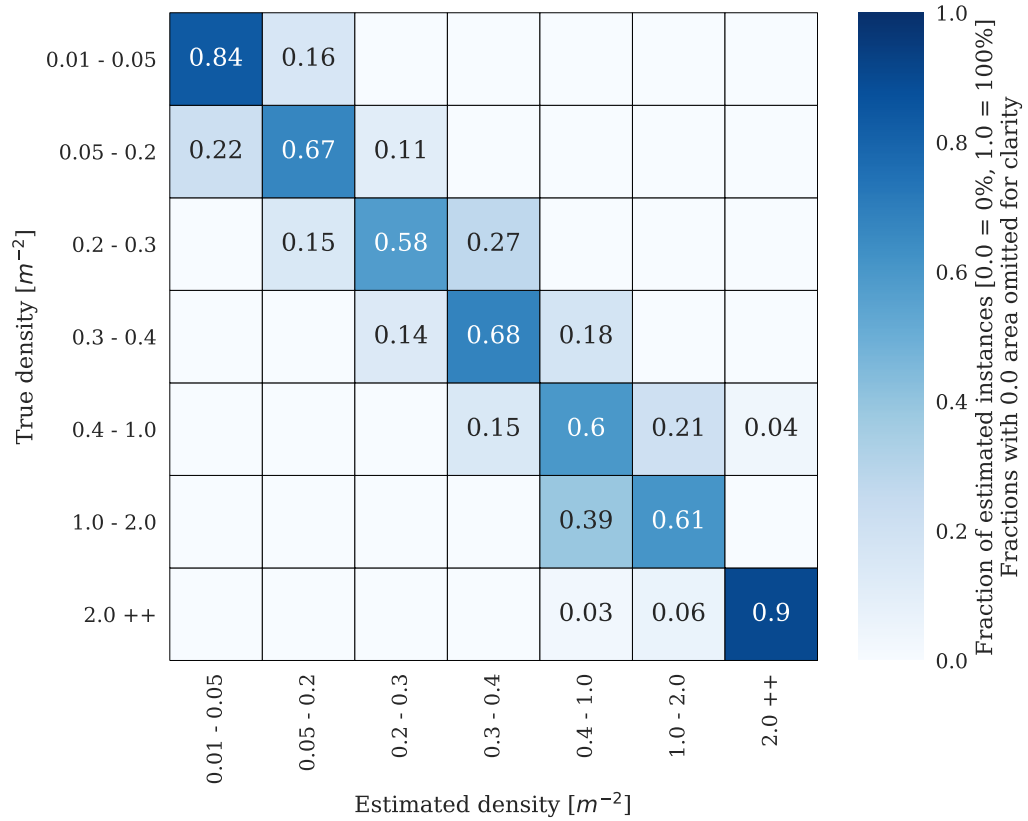
features for estimating the correct crowd density class on seven discrete crowd density classes with stratified 10-fold cross-validation (see Table 4.5 on page 129). This outperforms the reference method of absolute individual features by 21 % (reference method score 56 %). In the confusion matrix we observe that the majority of the predictions are distributed along the diagonal -with few exceptions. This means mis-classifications just between directly adjacent classes to the true crowd density class. Only classes ‘0.4 - 1.0’ and ‘2.0 ++’ are mis-classified just 3–6 % of the instances to not similar classes. The different crowd density classes are classified with a classification score between 0.56 and 0.94. Figure 4.20 on page 132 shows the confusion matrix of the collaborative absolute features.

**Table 4.5.** Crowd density estimation classification results based on the public viewing experiment validation. Includes the results of the collaborative approach with different methods. Includes the results of the reference method of individual sensing.

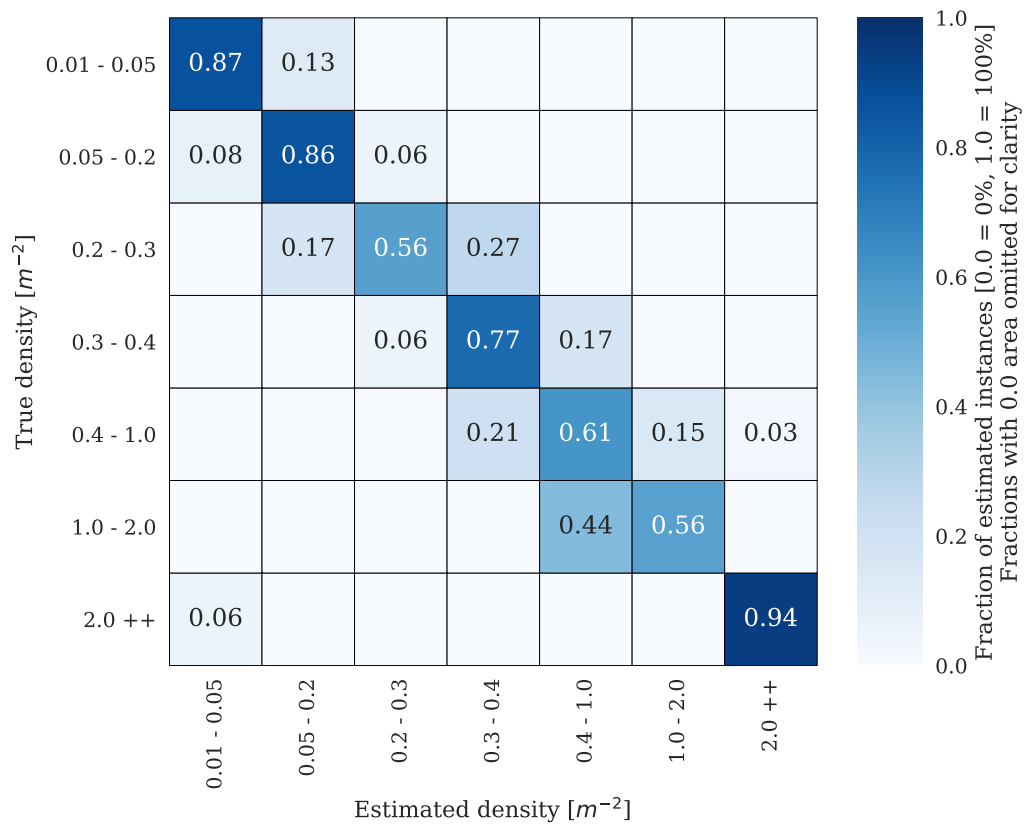
Feature Set	Sensors	Classification results			
		Score	Precision	Recall	F1-Score
collaborative absolute	10	0.67	0.71	0.69	0.66
collaborative relative	10	0.74	0.80	0.78	0.77
collaborative absolute&relative	10	0.77	0.83	0.74	0.75
individual absolute (reference)	10	0.40	0.36	0.38	0.38
individual relative (reference)	10	0.57	0.62	0.57	0.51
individual absolute&relative (reference)	10	0.56	0.63	0.56	0.54



**Figure 4.18.** Confusion matrix of machine learning stratified cross-validated classification results based on *absolute* features.



**Figure 4.19.** Confusion matrix of machine learning stratified cross-validated classification results based on *relative* features.



**Figure 4.20.** Confusion matrix of machine learning stratified cross-validated classification results based on *all* features including absolute collaborative features and relative collaborative features

## 4.8 Conclusion

This thesis chapter demonstrated that with just a few scanners and a low saturation of discoverable crowd devices (5–7% of people having Bluetooth enabled) we can achieve a viable mobile ad-hoc based crowd density estimation. We presented two general methods based on a single group of people with scanning mobiles and multiple collaborative groups of people with scanning mobiles being intermittently within the proximity.

We presented a method to rely not just on the number of devices seen by a scan, but also take into account information about average observed signal strength and the variance in both the signal strength and the number of devices. This makes the system more robust against variations in the number of discoverable devices that may result from the background of the people in the crowd rather than the crowd density. We demonstrated that relative features based on data collected with a group of synchronous scanners allow a crowd density estimation performance score of 0.55 with mis-classifications in just neighbored classes. We investigated the benefit of combining the information from several devices carried by different close by users, rather than on an individual scanner. When combining absolute and relative features based on the group scanning we achieve a viable crowd density estimation precision of 0.66 with mis-classifications in adjacent crowd density classes.

The just over 66% accuracy on four classes must be seen in the context of noisy ground truth resulting from arbitrary class definition, extrapolation between photos taken every 500 meters, and inaccuracies in the counting process. In addition, confusions occur nearly exclusively between adjacent classes (see Figure 4.6 on page 110 and Figure 4.7 on 112). Note that the experimental data did not include the ‘nearly empty space’ class which can be trivially recognized from the near absence of Bluetooth devices and could be easily integrated into the system.

We presented a method to combine the collaborative sensor information from several mobile phones carried by different groups of static and dynamic intermittently close by users. The core of the method is the comparison and fusion of data from different devices which leads to up to 27% improvement in accuracy over a simple single device(s) approach. We presented that the collaborative relative methods

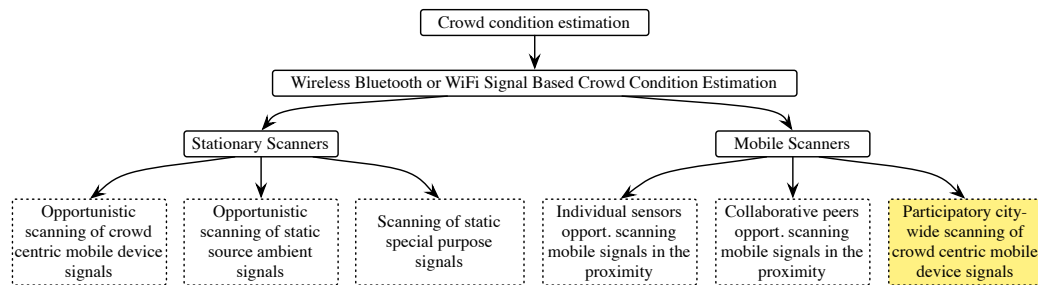


are exceeding the results from the collaborative absolute methods - building rather on the signal variations than on the background of the people. The just about 77% accuracy on seven classes must be seen in the context of noisy ground truth resulting from arbitrary class definition, extrapolation between the ground-truth based crowd density extraction every 10 minutes and inaccuracies in the counting process. In addition, confusions occur nearly exclusively between adjacent classes (see confusion matrices in Figure 4.18, and Figure 4.19, and Figure 4.20). Note that the experimental data did not include the ‘totally empty space’ class which can be trivially recognized from the near absence of Bluetooth devices and could be easily integrated into the system.

# 5

## Participatory Citywide Sensing

Jens Weppner, Paul Lukowicz, Ulf Blanke, and Gerhard Tröster. Participatory bluetooth scans serving as urban crowd probes. *Sensors Journal, IEEE*, 14(12): 4196–4206, Dec 2014

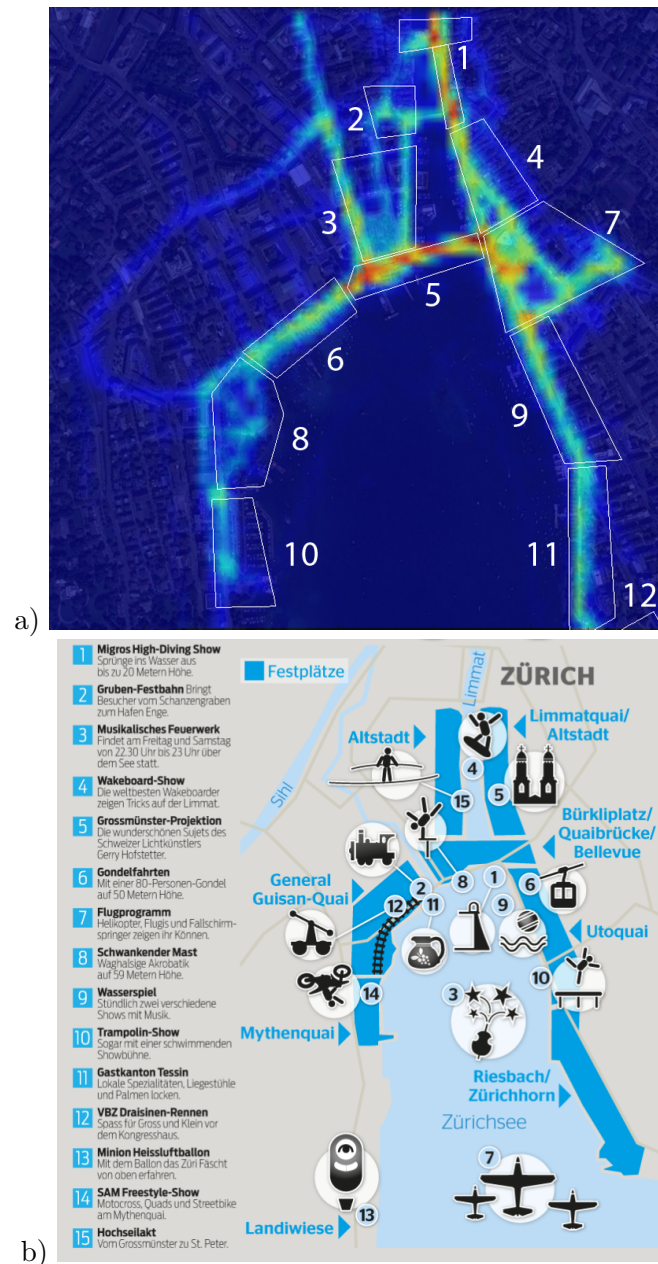


**Figure 5.1.** Thesis outline and wireless signal based crowd condition estimation scanning ontology.

## 5.1 Introduction

Crowds are an integral component of urban environments: from city festivals through sports events to rush hour in busy business or shopping districts. As a consequence monitoring, managing and planning for crowds is a key concern of civil protection and city authorities. Today the main instrument of crowd monitoring are CCTV cameras. While useful in many situations, they are, however, more suitable for intensive surveillance of constrained hot spots during well defined time periods than for long term monitoring of large areas. Alternatives (see related work) that have been considered range from airborne cameras through cell tower information to counting people at access control points (where possible).

As another alternative related work along the lines is investigating smartphone based participatory approaches. The core idea is that smartphone apps are increasingly becoming basic tools of daily city life. This includes navigation, public transport (including online tickets), information about services and opening hours, tourism and special events. In particular, large events such as city festivals are today unthinkable without an own App. Our work leverages such apps asking users to voluntarily contribute data for crowd monitoring. Originally, related work had focused on anonymized location information and estimating crowd density distribution from the distribution of data points provided by the volunteers. In a trivial approach one can simply count the number of people providing data from a certain location, assuming that they constitute a fixed percentage of the crowd, and then extrapolate to the number of people present. In reality, the procedure is more complex (as the percentage may be neither known nor constant), however, related work has shown that, given enough participants, a good estimation of the crowd density as well as other parameters such as speed, flow direction is indeed possible [105]. A major concern observed was how to achieve sufficient participation. Thus, for example, getting a few hundred to a thousand participants for large scale city events was not a problem (in fact, these could come from the organizers and civil protection forces). Getting tens of thousands participants is also possible as we had around 55 000 app downloads for the event App in Zurich (Switzerland) which is the foundation for this chapter. However, it requires an extremely well prepared, very intensive marketing campaign that is often not feasible.



**Figure 5.2.** Experiment and event visualization with a) the crowd density heat-map snapshot on Friday (05.07.2013, 9 p.m.) based on GPS location data transmitted to the server during 60 minutes. The main event areas during the *Zürifäscht* 2013 event are shown as polygons with white borders. b) The pre-determined event areas by the organization authorities of the event (image source: [www.blick.ch](http://www.blick.ch))

5 Participatory Citywide Sensing



**Figure 5.3.** Impression of the 'Zuerifaescht' event in Zurich (Switzerland) in the year 2013. (Image sources: a) [www.20min.ch](http://www.20min.ch) and b) [www.srf.ch](http://www.srf.ch))

## 5.2 Motivation and Problem Statement

In summary, the question is how participatory crowd monitoring can be extended to situations where the number of participants is too low to represent the crowd distribution and motion in a statistically significant way from their GPS traces alone (in other words the participants are too sparsely distributed within the crowd to accurately reflect its structure and motion). The proposed solution is based on the following observations:

1. Many users leave their smart phone bluetooth subsystem in discoverable mode ‘per default’ e.g. for the convenience of just getting into the car and being automatically connected to the speakerphone (see Figure 5.5 on page 149).
2. Scanning for discoverable Bluetooth devices is a standard functionality in most smartphones so that participants’ devices can be made to transmit not only their GPS data but also information on discoverable devices that are within their range.
3. In general the Bluetooth range is limited to 10 meters. This means that adding information about discoverable devices in reception range to participants’ GPS data is equivalent to providing location information not only about the participants but also about the owners of the discovered devices. This effectively increases the size of the sample that can be used for crowd density and motion estimation.

Previous studies (see chapter 4, section 4.6 and section 4.7) with students as participants carrying scanning smartphones

- at small scale events (thousands of people in an areas of about  $500 \times 20 \text{ m}$  and  $200 \text{ m} \times 200 \text{ m}$ ),
- following well defined walking patterns

have confirmed the basic feasibility of using such an approach to accurately estimate crowd density. However, they have also shown that the actual number of discoverable devices can vary strongly in space and time for a given crowd density so that only very rough estimates are possible when using the absolute number of discovered devices as a feature. Motion patterns of the scanning devices

have significant influence on the performance. Thus, the core scientific questions addressed in this chapter are defined by the following aspects.

1. *How does such an approach perform in unconstrained city scale environments where participants **are not** students following well defined motion patterns but ‘normal people going about their business’?*
2. *What does it take to improve the system performance under such conditions, in particular in terms of choosing and designing features that go beyond a mere device count?*
3. *Is it possible to go beyond density estimation towards the recognition of motion patterns even through the owners of the discovered devices (who do not actively participate in the data collection and do not provide GPS data) from whose have the approximate location but have no motion information?*

### 5.3 Chapter Overview and Contributions

We describe a system that leverages users voluntarily having their smartphones scan the environment for discoverable Bluetooth devices to analyze crowd conditions in urban environments. Our method goes beyond mere counting of discoverable devices towards a set of more complex, robust features. We also show how to extend the analysis from crowd density to crowd flow direction. We evaluate our methods on a data set consisting of nearly 200 000 discoveries from nearly 1000 scanning devices recorded during a three day citywide festival in Zurich. The data set also includes as ground truth 23 million GPS location points from nearly 30 000 users.

Towards answering the above questions the chapter makes the following contributions:

1. A large, real life data set with nearly 1000 devices (subsequently called *Bluetooth scanner* or *scanner*) providing Bluetooth scans (nearly 200 000 discoveries) annotated with location information over a period of three days during a citywide festival in Zurich. The data set also contains the ground truth for the density and motion analysis that is based on around 30 000 users providing their GPS coordinates.

2. Use of the data set to evaluate the naive crowd density estimation method (extrapolating from the number of seen devices) against the GPS based ground truth.
3. A more advanced method that goes beyond absolute numbers towards relative features that are more robust against statistical variations of the number of devices present at a given density. The method is evaluated on the same data set and compared to the naive method.
4. A method for the estimation of the crowd flow direction, again with evaluation on the data set against the GPS based ground truth.

## 5.4 Related Work

Our work deals with (1) participative (2) crowd state analysis estimation using (3) Bluetooth scanning. The relevant state of the art research in the three areas can be described as follows.

### 5.4.1 Participatory Sensing

Among others Campbell et al. [16] and Burke et al. [15] introduced the general concept of people-centric sensing and participatory sensing. Since then a lot of work has been done in this area including sound pollution [85], air pollution [29] or road and traffic conditions [68]. Wakamiya et al. studied temporal patterns of crowd behavior indirectly speculated from a massive number of collected Twitter messages [95]. In previous work [105] it was demonstrated how participatory collections of GPS traces can be used to monitor the crowd condition (this is being used as ground truth for the Bluetooth methods described in this chapter).

### 5.4.2 Crowd Monitoring

Video based crowd analysis became popular in the 1990s with the increased use of CCTV cameras and availability of sufficient computing power (e.g. [27]). Since then extensive research has been done and a comprehensive overview goes beyond the scope of this chapter (see. e.g. [111]). Examples of specific work range



from detection of anomalies in crowd behavior [65], through work related to privacy preserving analysis [18] (not tracking or identifying individuals) to various multi-camera systems [91]. Significant attention has also been given to tracking individuals in crowds [54] including large area tracking with multiple cameras [50]. Overall the video monitoring work must be seen as complementary rather than an alternative to our research of long term large area participatory analysis being complemented by punctual video surveillance of specific hotspot. An alternative may be airborne cameras that can cover large areas [86].

Beyond camera-based crowd monitoring, methods based on thermal imaging [2], combination of thermal imaging and cameras, [5], wireless sensor network signal propagation [109], cell tower information [79], and passive RFID monitoring [98] were proposed.

### 5.4.3 Bluetooth Scanning

With the proliferation of mobile Bluetooth enabled devices leveraging the information about discoverable devices has become an active research field in Ubiquitous Computing. Early well known work showed [30, 40] how to recognize social patterns in daily user activity, infer relationships and identify socially significant locations from using Bluetooth scans. Since then Bluetooth has been widely investigated as an additional source of information for various activity and lifestyle monitoring systems (e.g. [34]).

Towards public spaces and crowd related applications Nicolai et al. [75] looked at the discovery time of Bluetooth devices and the relation between the number of people and the number of discoverable Bluetooth devices. However, unlike in our approach only the absolute number of discovered Bluetooth devices was used. Morrison et al. [70] considered the visualization of crowd density in stadium-based sporting events. In [53] the authors recorded passenger journeys in public transportation by analyzing Bluetooth fingerprints. O’Neill et al. [76] presented initial findings in Bluetooth presence and Bluetooth naming practices. Versichele et al. [92] performed an experiment during a mass event where they covered an area with static Bluetooth scanning devices to extract statistics and visitor profiles. BLIP Systems [12] exploited a stationary Bluetooth based people tracking system.

Based on multiple Bluetooth zones scenarios like queue length at airports or travel times by car are indicated.

With respect to large scale applications of mobile Bluetooth sensing Natarajan et al. [73] have had 12 participants scan a city for discoverable Bluetooth devices over a period of three months. A similar study was conducted on a larger scale (100 devices, nine months) by Henderson et al. [39]. Finally, there were different studies with the scope of university campuses and conference locations (e.g. [89]). Overall the work we present differs from those in this scope (we base our work on an order of a magnitude of more of Bluetooth scanning devices and discoveries) and we focus on crowd behavior analysis. In our previous and initial work we have demonstrated the feasibility of an early version of the features described in this work at a small scale experiment with instructed students [100].

## 5.5 Data Set

The data set that this work is based on has been recorded during a three day citywide festival in Zurich (Switzerland) in the summer of 2013 (<http://www.zuerifaescht.ch>). The festival takes place every four years and attracts up to 2 million people with a mixture of shows, concerts, sports events, parades and parties distributed all over the city. The recording had been leveraged by the event management platform developed during the Socionical European Union project (<http://www.socionical.eu>) and tested (mostly at a smaller scale and without Bluetooth scanning) at a variety of events in London, Zurich, Vienna and Amsterdam. The platform is build around an event information App [11] which the attendees can use to plan their visit and get information on anything from the location and timing of events through the background of the festival to public transport and route planning. The app also includes a variety of social networking features. In parallel, it integrates a set of safety/security modules which the users could activate on a voluntary basis:

1. A monitoring module that records and transmits data of a set of selected sensors to the server. The sensors are requested once the app was launched for the first time and require explicit user consent for every sensor.

2. A location sensitive messaging module that allows the organizers to send information or instructions to users at specific locations or at users heading in specific directions.
3. A privileged module that is activated via special code when the app is not being used by a visitor but by a member of the civil protection forces.

Considering the event management the collection of anonymized GPS traces, their visualization in form of a heat map and the location based messaging capability were the key. For the experimental purposes described in this thesis chapter, for the first time of this software platform, users were asked to activate their Bluetooth module if it was not activated previously and if they agree with scanning for Bluetooth devices even when the application is currently not used. Synchronously to collecting the data users were asked to transmit the Bluetooth discovery information, together with signal strength, identifier and timestamp. The Bluetooth data collection procedure had been previously cleared with the Zurich legal authorities.

### **5.5.1 Experiment Advertising Campaign and Distribution**

We endeavored to achieve a very high quantity of participators acting as urban crowd probes. There are primarily three goals to achieve. Getting the users to download the application and acquire the permission from the user to collect the sensor data in compliance with the privacy policy (see sub-section 5.5.3 for details). The first goal was successfully achieved (55 000 app downloads) by collaborating with the event management, local media featuring the scientific crowd sensing aspects of the event application. Substantive functionality such as the schedule and site information of the festival were of high interest by the users. Once the potential users were aware of the application the users downloaded it via the Apple and Google app stores. Most importantly, collecting sensor data (GPS localization and Bluetooth scans) while in the event area in the background must happen with clear communication with the user i.e. why, when and in which area sensor data is collected. To let users easily participate in collecting sensor data no explicit registration was necessary. As a result of the advertising campaign 55 000 people

downloaded the application and a total of 30 000 people (approximately 54% of the app downloads) uploaded sensor data to the server. Users not participating either opted-out, deactivated their data uplink, or never initially launched the application after downloading. Application support was built into the application giving hints on how to use the app, the privacy policy, and the possibility how to opt-out regarding the data collection and transmission process.

### 5.5.2 Privacy Policy and Anonymization Approach

Most importantly, data protection officers made clear to precisely communicate that data is used and how it is used. Through press releases the public has been completely aware of this experiment. While the user initially launched the application an indication about scientific and safety rationales behind the data collection and data transfer to the server was shown along with a guide on how to opt-out. This had to be confirmed by the user prior to any usage of the application, data collection or data transfer. No information was transmitted to the server which would infer to an identity of a participant. We emphasized not to annotate data transfers to the server with any permanent user name or smartphone identifier (device MAC address, device UUID, etc.). Additionally the GPS localization mechanism had to be accepted due to operating system requirements for newly installed applications (on iOS at the first launch of the app, on Android requested device permissions are displayed prior to the download of the app). After the confirmation a temporary *random event device identifier* was generated which was sent together with the GPS and Bluetooth data packets to the server. The *random event device identifier* cannot be mapped to a user identity and has the life time of the special purpose event application. Any conclusion of anonymous traces of *event device identifiers* is not possible since location recording and transmission is limited to the  $1.5 \text{ km}^2$  event area and prominent user locations like the beginning and end of a trace (i.e. location of residence, location of work, etc.) were not collected. The IP address of the device (incoming data packet sender) was not stored. Next to the anonymization of the participator we considered the anonymization of Bluetooth discoveries. Each Bluetooth discovery contains a MAC address which is uniquely assigned during the manufacturing of the Bluetooth chipset. We uploaded

the MAC address to the server where we used a random salt which was used as an additional input to a one-way hashing function (SHA) to encrypt the MAC addresses irreversibly.

### 5.5.3 Experiment Procedure and Data Collection Process

The distribution to a wide audience is more complex than to a small set of persons. As we had to distribute the app through the official app stores, certain technical and regulatory requirements had to be met. The team responsible for the app and its release consisted of three persons. According to the app store carrier Apple, the regulations of a background process accessing the GPS location was not allowed without any direct benefit to the user. For this reason a location-based feature called ‘friend finder’ was integrated into the application for getting Apple’s app store approval. After the app download, the initial launch and the privacy policy acknowledgement the application configured itself to start recording experimental data for scientific research in the morning of the first event day. A data packet was sent every two minutes (or buffered in case of 3G network congestion) to the server (4 Amazon AWS server instances) running MongoDB data base instances. When exiting the event zone the GPS localization was switched off. In the night of the last event day the data collection module was deactivated automatically to prevent collecting and uploading of unintended data in case the user kept the application on his device. The experiment logic was integrated into the application. The operating system function called ‘geo-fencing’ (coarse but power efficient location method based on cell tower locations) automatically activated the data recording process in the background if the user was present in the event area which covered  $1.5 \text{ km}^2$ . When data recording was activated GPS data was acquired at 1 Hz, and Bluetooth scans were obtaining every minute. The core part of the experiment was the collection of Bluetooth scan information. Bluetooth scanning is defined as the process of recognizing surrounding Bluetooth devices. Each Bluetooth scan can result in  $n \geq 0$  Bluetooth discoveries. Each discovery contains information of the device name (ignored), device profile (ignored), supported services (ignored), unique MAC address, timestamp and signal strength. The duration of a Bluetooth scan

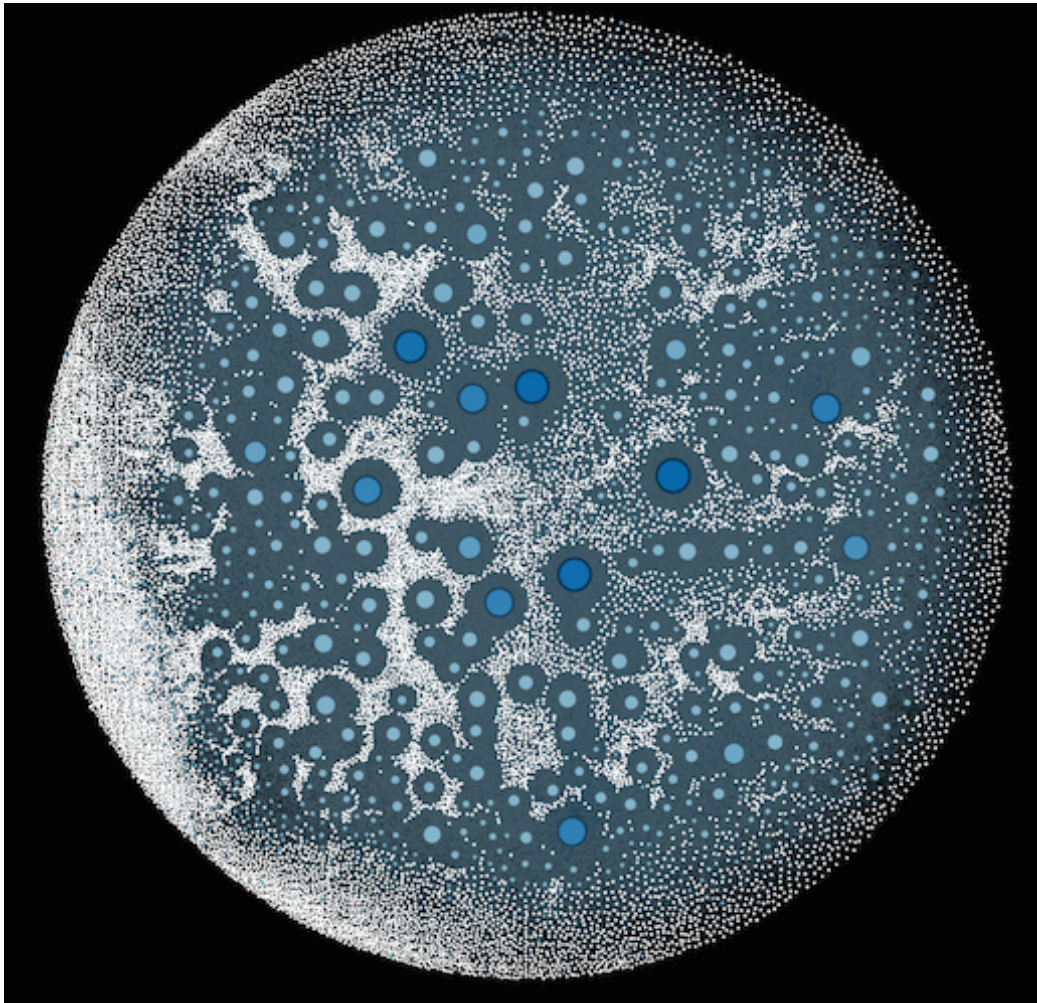
(as of current Bluetooth chipset and operating system cooperation) is dynamically controlled depending on whether new devices (within the scan period) are detected. This is motivated by energy saving of the Bluetooth module when no devices are discovered. The data was stored on the server for offline analysis.

#### 5.5.4 Data Characteristics

Some key statistics of the collected ground truth data and Bluetooth discovery data are shown in Tables 5.1. From the about 2 million visitors 55 000 had downloaded our app and 30 000 of those have been actively transmitting GPS data. Of those 971 have also provided Bluetooth scans. This is due to the fact that users had to explicitly activate the Bluetooth module and many were worrying about power consumption issues or simply shunning the effort. Over the course of the event this gave us nearly 200 000 discoveries that belonged to around 20 000 unique devices.

#### Distribution of Bluetooth Discoveries

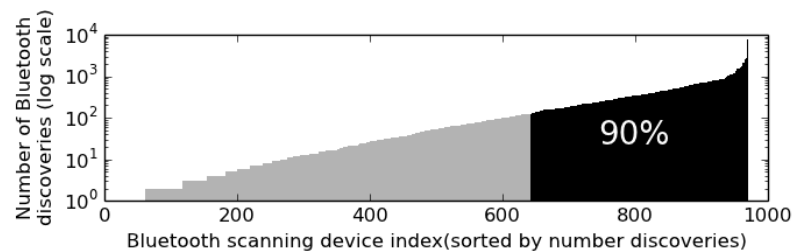
The vast majority of scans has turned up relatively few devices. Figure 5.6 shows a comparison of the statistics from Zurich to five other events: two football games (at ‘Wembley’ stadium in London and at the ‘Allianz Arena’ in Munich), the Munich October Fest, a festival in the city center of Valetta in Malta (very small area compared to Zurich) and the public viewing soccer event in the German city of Kaiserslautern. What all the other events have in common is that a small number (10) of Bluetooth scanners were moving around a constrained, very crowded area. Thus the majority of Bluetooth scanning periods turned up a value corresponding to the typical number of discoverable devices in a dense crowd which was somewhere between 5 and 20 depending on the crowd and the location. It is also interesting to note the similarity in the shape of the distribution of the Zurich event, the Malta festival and the Allianz Arena data. The three had a comparatively larger area going beyond a single crowded location (in the Allianz Arena experiment the data was collected around rather than inside the stadium). However, the Zurich distribution is much more distinct, due to the much larger area.



**Figure 5.4.** The aggregated Bluetooth topology of the whole event duration visualized as a graph. The circular layout is caused by the spring-embedder based ForceAtlas2 (gravity and repulsion based) graph visualization algorithm. Due to the large number of edges (an edge equals a distinct Bluetooth discovery) these appear as bluish blur in the background. Blue dots represent Bluetooth scanners. The larger a blue circle (and proportionally more blue) is the more discoveries were made by a certain device. Small white circles represent discoveries.

## Proportion of Relevant Bluetooth Scanning Devices

The vast majority of Bluetooth discoveries comes from a relatively small number of devices. This is illustrated in Figure 5.5. Exactly 329 devices were accountable for 90% of the total number of discovered Bluetooth devices. Figure 5.4 visualizes those devices as large blue circles and discoveries as small white dots. Discoveries are arranged in the proximity of the device(s). Again, the nature of the event explains the data. Many people would visit the event briefly or stroll through the city streets rather than spending more of their time at crowded locations. Additionally, a number of participants had the Bluetooth scanning functionality turned on only briefly.



**Figure 5.5.** Distribution showing the proportion of Bluetooth scanning devices to be accounted for Bluetooth discoveries. The bar chart shows the index of the Bluetooth scanning devices in sorted order (x-axis) with respect to the number of individual Bluetooth discoveries (y-axis). The y-axis is shown in *log* scale to visualize the wide range of discoveries from one to 7964 Bluetooth discoveries per Bluetooth scanner. Apparently, broad bars are not to be confused with a single bar but multiple bars close to each other.

## Uniform Event Area

Of the nearly 1000 scanning devices only 13 have seen each other over the course of the festival. Given the large temporal and spatial extent of the festival and the fact that the scanners were a random selection of the participants this is not surprising. People were at different places at different times. What is surprising is the fact that over 700 scanners shared at least one device that they have both discovered over the course of the festival. In fact more than half of the scanners shared at least 20 devices. This implies that the festival did not strongly separate



into distinct events with little shared spectators. The above points illustrate that participatory Bluetooth sensing is not only suitable for assessing crowd density (as is the focus of this chapter) and flow directions but that it contains information about more complex aspects of an event and may be used to recognize different types of events taking place in a city.

**Table 5.1.** Data set statistics of the mobile participatory wireless scanning experiment in Zurich.

<b>Event Attribute</b>	<b>Value</b>
Event duration	3 consecutive days
Scope	1.5 $km^2$ event area
Estimated number of visitors during the event (according to event organizers)	2 million
Ground truth entities (total number of GPS locations collected and uploaded)	23 million
Number of app downloads	39 300 (iOS) + 15 600 (Android OS)
Devices collecting and uploading GPS traces	23 400 (iOS) + 6400 (Android OS)
Average number of location samples per device	Friday: 586 Saturday: 643 Sunday: 703
Average time collecting GPS locations (including pauses)	Friday: 12 840 seconds Saturday: 14 378 seconds Sunday: 10 145 seconds
Users actively participating in collecting and uploading Bluetooth scan data (Android OS)	971
Total Bluetooth discoveries	190 600
Distinct Bluetooth discoveries	18 900

## 5.6 Citywide Area based Crowd Density Estimation

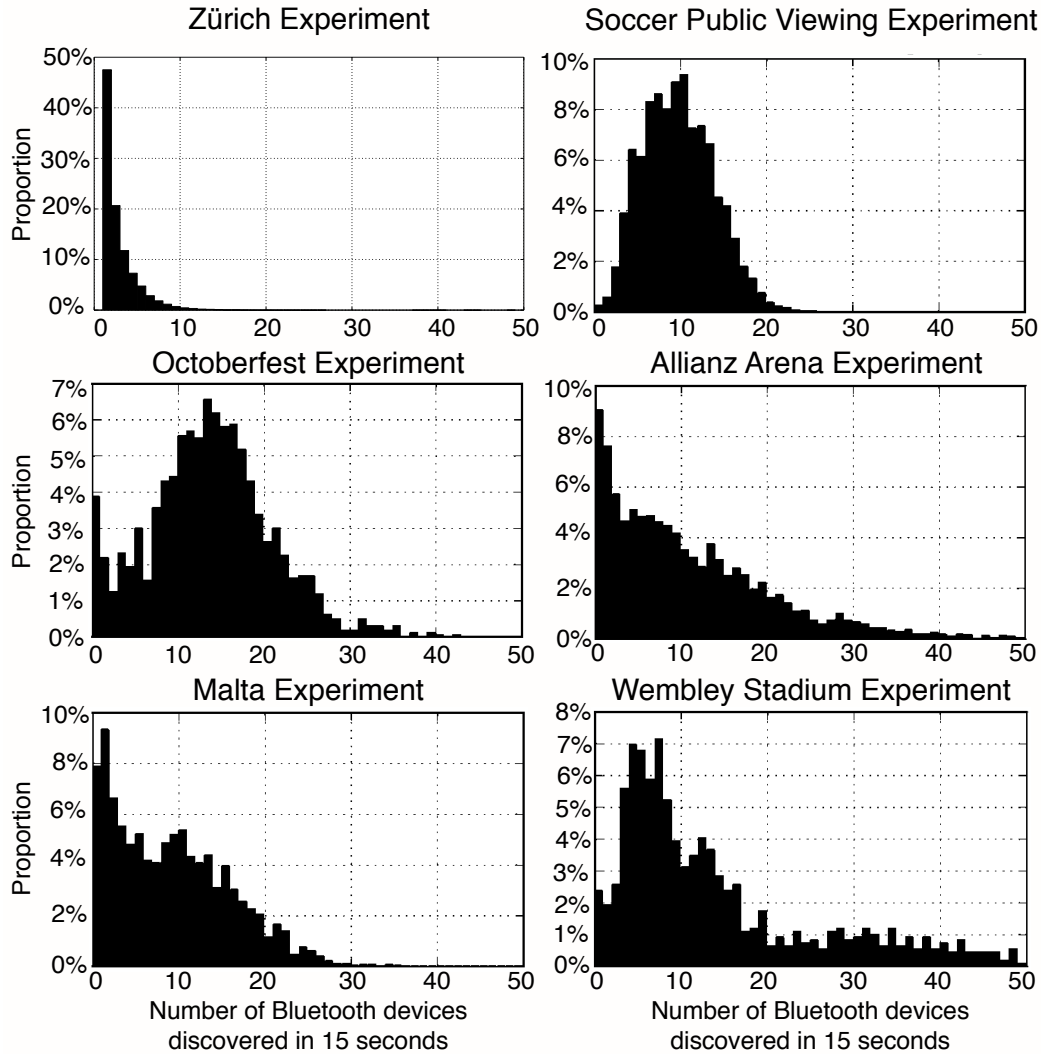
We analyzed the crowd condition estimation within defined areas during the Zurich event. A typical area has the size of thousand to tens of thousands of square meters. Areas are defined manually as being thematically associated. People are attracted by an area, by its certain topic such as drink stands, open-air music concerts, open air shows, spectator zones along the Zurich lake front for on-water shows, activity zones for children and adults, public transit zones and pedestrian passage zones between areas. A small number of Bluetooth scanning participants can deliver fast crowd condition results within an area representing the current crowd condition. An advantage of area based monitoring is the low latency estimation because scanning just small parts of the area allows an assertion of the whole semantically connected area. For a basic crowd condition estimation not every angle of the area needs to be monitored before presenting a first estimation. Another advantage is the clear connection between the crowd condition and the affected area. Security staff monitoring the crowd condition estimation in real-time can easily recognize the potentially affected area. In total 12 pre-defined areas mapped to activity zones were monitored.

### 5.6.1 General Principle

An obvious way to estimate the crowd density is to perform a scan for discoverable devices and assume that the number is an indication for the number of people in the vicinity defined by the Bluetooth range (typically around 10 meters). Unfortunately, this simple approach contains a number of problems. Firstly, there is the issue of sufficient statistics. With the scan limited to a radius of about 10 meters (approximately a circle with an area of  $300 \text{ m}^2$ ) anything between a few and a few hundred people can be within range. While in a dense crowd with a few hundred people we may get a representative sample, in less crowded areas we are likely to see very strong variations between samples. Assuming the probability of any single user having a discoverable Bluetooth device to be in average 10% and 20 people are within range the average number of discovered Bluetooth devices will only be two. While meeting in a space with exactly two technophile friends having

Bluetooth switched on will also result in two Bluetooth discoveries. Thus we may sometimes be in a group of people who do not even have activated mobile phones while at other times we may be surrounded by a group where everyone has an active Bluetooth device. Secondly, there is the question of signal attenuation. At 2.4 GHz (which is the transmission frequency of Bluetooth) the human body has a high absorption coefficient. This means that in a dense crowd (where we would expect to have good statistics) the effective scan range is reduced and therefore ‘falsifying’ the results. Finally, we have to consider cultural factors. This means that the average number of people carrying a discoverable Bluetooth device may significantly vary depending on who the persons in the crowd are. For the same crowd density at a student party of a technical university a different number of devices may be present than at a fifth division soccer game in a poor rural area. In the previous chapter 4 we investigated a group-wise and collaborative Bluetooth based crowd density measurement approach. However, the methodology of the previous experiments was different and served as an initial study on the feasibility of Bluetooth based crowd density. In the previous work we performed an experiment in a controlled environment during a public viewing event during the European soccer championship (see Figure 5.6 on page 153, a soccer public viewing event in Kaiserslautern, Germany). The experiment persisted of 4 hours (1 hour during arrival, 2 hours during and 1 hour during departure). The experiment took place in a rectangularly fenced area with just a single entry and exit point while most of the event visitors stood on the spot without moving after the arrival. During the experiment we instructed students in five groups each of two people to move consistently along a pre-defined imaginary path within the fenced area during the three periods of the experiment. A walk along the path was finished in less than three minutes and then repeated in reverse. In the previous approach we developed features and built the method on Bluetooth discovery of constantly two nearby scanning devices (one device per person, two per group) where we analyzed variations of Bluetooth discoveries between both participants with a fixed spatial connection.

In the work described in this chapter we have not set any requirements to the participants behavior. Nearly 1000 participants moved freely at any desired speed and direction without any influences from us. We applied neither methods from



**Figure 5.6.** Visualization of Bluetooth discoveries in relation to a consistent 15 second time windows. The x-axis of each sub-plot represents the number of Bluetooth discoveries and the y-axis represents the proportion (percentage) of time windows with a certain number of Bluetooth discoveries. The time windows are based on the whole experiment duration and on all experiment participants. Each sub-plot represents one experiment.

previous work nor analyzed continuous groups of participants since time periods of two constant close-by Bluetooth scanning participants was insignificant. The covered experiment area is heterogenous consisting of many streets, footways,

pedestrian zones, parks, stages and food courts which are divided by buildings, bridges, a river and a lake. Visitors either stood statically at one point (i.e. at a music stage, at the water-front spectator zone, etc.) or moved in the same direction (i.e. before beginnings of mass events) or moved in different directions while strolling around or spreading out (i.e. from train stations, etc.).

### 5.6.2 Advanced Method

The proposed advanced method builds on features which go beyond of just counting Bluetooth discoveries. Our main contribution lies within the new features presented in subsection 5.6.2. Bluetooth scan information is the main component of the feature set but also GPS sensor information is taken into account. Our approach was to aggregate sensor data from multiple participants to obtain a statistical validity which had the aim to achieve a higher robustness regarding noise and estimation accuracy compared to the trivial approach by just counting the Bluetooth discoveries. This aggregation was applied to twelve different event zones defined by us according to the event schedule and event map. Secondly, we aggregated the sensor data by time, either with a time window of 10 or 30 minutes. As a result of the spatio-temporal aggregation we obtained one 12-dimensional feature vector per time window and event area. All in all we obtained 5184 (for a 10 minute time window) respectively 1728 (for a 30 minute time window) feature vectors. For our regression analysis we built on top of established methods. We applied a feature selection using M5's method (step through the features removing the one with the smallest standardized coefficient until no improvement is observed in the estimate of the error given by Akaihe information criterion) and eliminated collinear features. After we obtained the set of feature vectors we built a regression model based on the feature vectors and computed the ground truth value for each feature vector. This crowd density estimation method was then applied on the feature set for evaluation.

## Feature Definition

We introduce 12 features which were computed based on the Bluetooth discoveries and the GPS location information of the nearly 1000 Bluetooth scanning participants. The data was stored in the MongoDB database and features were developed later after the end of the event. Information from other participants which served as a source of ground-truth was intended not to be involved in computing the features. In the following, the proposed features are described. (1) The *average speed* of the scanning devices (sensors) indicates different crowd states: If the speed is low either the scanner is stationary and the crowd is stationary (due to high crowd density or on-going event) or the scanner is stationary (spending time at food or drink stand) and the crowd is passing by. We calculated the *average speed* by averaging the speed values of all sensors during discovering Bluetooth devices. We did not take the speed of other non-Bluetooth scanning devices into account. (2) The *average Bluetooth signal strength* (RSSI value) reflects a rough statement about the average distance and signal attenuation between the scanning device and discovered devices. We calculated the average signal strength of all Bluetooth identifiers including multiple discoveries of the same Bluetooth identifier by one or multiple scanners and then averaged the value. (3) The *variance of the Bluetooth signal strengths* indicate the deviation (due to different distances and signal attenuation) of signal strengths. We calculated the variance of the signal strength of all Bluetooth discoveries including multiple discoveries of the same Bluetooth identifier. (4) The *variance of subsequently measured Bluetooth signal strengths* for a specific Bluetooth identifier is influenced by the crowd behavior. If two devices are in the same distance to each other and does the crowd not move in between the signal link, the variance is lower than in a moving crowd. We calculated the variance of the signal strength of Bluetooth discoveries with the same Bluetooth identifier detected by the same sensor. We then averaged the values of all discoveries and sensors. (5) The feature is defined as the *average value of re-discoveries of each Bluetooth identifier address* by all sensors representing the overall crowd motion in an event area. We calculated this by counting the re-discoveries of the same Bluetooth identifier by any sensor. All sensors act as an aggregated sensor, as if the discoveries were coming from one sensor. If a re-discovery was made by the same sensor or another sensor is

irrelevant. We then average the number of re-discoveries over all current Bluetooth identifiers. (6) By analyzing the *average number of scanners discovering a certain Bluetooth identifier* per time window we can make an assumption of the coverage and scanner distribution in the event area. We calculated this feature by the sum of sensors which discovered a unique Bluetooth identifier and then averaged this over all current Bluetooth identifiers. (7) The *diversity of individual sensors* is defined by the overall average of the relation of uniquely discovered devices to the sum of non-unique devices discovered. This feature was calculated by the sum of Bluetooth identifiers which were only discovered by one sensor, divided by the sum of Bluetooth identifiers discovered by two or more sensors. (8) We define the *duration of device visibility periods* in a given time window as the maximum timespan a Bluetooth identifier was recognized by all sensors. Averaged over all scanners the length of the stay depicts the potential to be discovered by any sensor in the area. This feature was calculated by retrieving the first and last occurrence of a Bluetooth identifier which was discovered by any sensor in the time window. We then averaged the duration of all Bluetooth identifiers. (9) The *average time of Bluetooth sensors in the area* measures the ‘scan-ability’ of an area and takes time spans into account where no or few Bluetooth devices were found. We averaged the duration of active sensors in the given time window and, of course, the given area. We include three basic features in our feature set. (10) The *total number bluetooth discoveries* reflects the sum of all Bluetooth discoveries including re-discoveries of the same Bluetooth identifier by any sensor. (11) The *unique Bluetooth device discoveries* reflects the sum of all Bluetooth discoveries excluding re-discoveries of the same Bluetooth identifier by any sensor. (12) The *number of active scanners* is another measure of the ‘scan-ability’ of a certain time window and event area. This is calculated by the sum of all active sensors in the given time window and of course the given area.

## Ground Truth Definition

To evaluate our proposed Bluetooth based crowd density estimation method we had to consider a comparison with the actual number of people in a certain area. Manual methods for obtaining ground truth information with a granularity of

10 minutes for each of the 12 event areas were not feasible. We did not have the resources to deploy multiple persons all day long for multiple days at the wide spread event areas manually noting the number of people around. Even having the man-power it would be impossible to continuously count the number of constantly moving people in a complex area with multiple entry and exit points. For this reason we designed our experiment to collect additional ground truth information. While nearly 1000 participants were obtaining Bluetooth discoveries, nearly 30 000 participants (23 400 iOS and 6400 Android OS) obtained ground truth information with an average daily duration of 3.5 hours on Friday, 4.0 hours on Saturday, 2.8 hours on Sunday with potential pauses in between (we defined an event day from from 4 am to 4 am). The average number of samples per GPS trace was 586 samples on Friday, 643 on Saturday, and 703 on Sunday (see Table 5.1). We extracted the ground truth values from the collected data set by counting the unique *event device identifiers* in a certain time window and event area. While we are aware that our ground truth value is a value smaller than the real number of people (not all people present participate in the experiment with the provided apps) we assumed a constant factor to be multiplied with our ground truth values to achieve the real number of people. Since we are interested to evaluate our approach as a method to obtain the crowd density based on a small sample (971 participants vs 2 million event visitors) compared to a larger sample we consider calculating the calibration factor in future work.

### 5.6.3 Evaluation and Results

We applied a feature selection using M5's method and eliminated collinear features. We identified features (feature identifiers (2), (4), and (9) defined in subsection 5.6.2 on page 155) which did not contribute to information content of the feature vectors. We then evaluated our crowd density estimation method in multiple ways. Firstly, by comparing our new crowd density method (*advanced*) to two kinds of basic methods previously used in literature. The basic reference method is defined by simply counting unique Bluetooth device discoveries (*basic*), which counts multiple discoveries of the same Bluetooth identifier only once. For each method we generated an individual regression model which is based *on all event*



**Table 5.2.** Results (lower value is better) by evaluating the model with data subsets regarding different event areas. For generating the model all feature vectors of all event areas were selected, while each set of feature vectors of event areas was used as a test set. A 10 minutes time window was used for data aggregation. A previous methodology (*basic<sub>2</sub>*) for crowd density estimation was compared against our newly proposed methodology (*advanced*).

Event area	Approximation error (lower is better)	
	Basic <sub>2</sub> Method	Advanced Method
Area 1	75.6 %	67.4 %
Area 2	80.0 %	47.1 %
Area 3	77.2 %	78.9 %
Area 4	92.3 %	88.3 %
Area 5	61.4 %	47.7 %
Area 6	95.0 %	71.8 %
Area 7	79.8 %	69.7 %
Area 8	80.4 %	70.6 %
Area 9	87.8 %	80.0 %
Area 10	62.8 %	69.3 %
Area 11	99.2 %	52.5 %
Area 12	80.1 %	97.4 %

*areas* and a temporal aggregation of ten minutes. As described before we have 5184 feature vectors, while they either have the dimensionality of one (basic reference method) or twelve (the proposed advanced method).

As the evaluation metric we selected the relative approximation error, expressed as the percentage of our calculated value deviates from the absolute ground truth value. Method *basic* results to a relative approximation error of 57%. Our *advanced* method leads to a relative error of 47%. This is a decrease of 10% regarding the relative error compared to *basic* which denotes a significant improvement of our method. We visualized the individual error values of 5184 feature vectors with the *basic* method in Figure 5.7a and the error values of the *advanced* method in

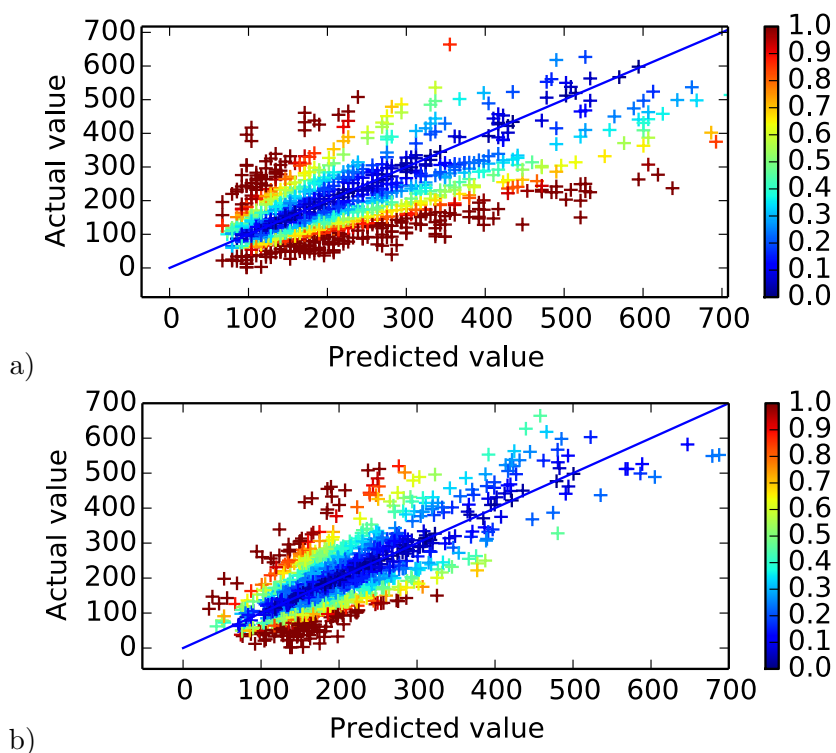
**Table 5.3.** Linear Regression results with data subsets. Showing the impact of the number of scanners ( $x$ ) involved to the crowd density estimation both on the correlation coefficient and relative absolute error.

Number ( $x$ ) of Scanners	Approximation error	
	10 min window	30 min window
$x \geq 2$	65.4 %	49.7 %
$x \geq 3$	66.4 %	50.9 %
$x \geq 4$	59.8 %	49.4 %
$x \geq 5$	52.5 %	45.7 %
$x \geq 6$	44.9 %	38.5 %
$x \geq 7$	39.8 %	33.0 %
$x \geq 8$	34.5 %	29.4 %
$x \geq 9$	35.7 %	27.5 %
$x \geq 10$	28.9 %	26.1 %

Figure 5.7b which show the deviations between actual value (x-axis, ground truth) and predicted value (y-axis, estimation) with a temporal aggregation of ten minutes. As visualized in the scatter plots the approach *basic* tends to exaggerate lower crowd density values to higher crowd density values, while our *advanced* method tends to concentrate values near to the diagonal line (representing 0% relative error).

Secondly, we evaluated the method on *individual* event areas with the proposed crowd density estimation method (*advanced*) and compared it to the reference method of simply counting unique Bluetooth devices (*basic*). With this evaluation we wanted to see whether the regression model, generated on  $n - 1$  event areas, fits to the  $n^{th}$  event area. We generated a regression model for all  $n - 1$  combinations while the set of feature vectors of the  $n^{th}$  event areas was used as a test set. A 10 minutes time window was used for data aggregation. The results are shown in Table 5.2. While our method outperforms the *basic* method in nine of twelve event areas, our method has a higher relative absolute error at two event areas and approximately the same relative absolute error at one event area.

Lastly, we analyzed the impact of the number of Bluetooth scanners involved in the crowd density estimation with the *advanced* method. For each method we generated an individual regression model which is based on *all* event areas and a temporal aggregation of 10 and 30 minutes. The raw data was not reduced by selecting a range of Bluetooth scanners on the data set but by selecting subsets of the feature vectors (each feature vector corresponds to a time-window and an event area) which complied to the given criteria. We filtered the feature vectors by the number ( $x$ ) of Bluetooth scanners actively scanning (not to be confused with the number of Bluetooth discoveries). Multiple subsets of the feature vectors with the attribute of  $x \geq 2$  up to  $x \geq 10$  Bluetooth scanners were selected. For the evaluation we used 10-fold cross-validation. The resulting relative approximation errors are shown in Table 5.3, which are ranging from 65% ( $x \geq 2$ ) to 28% ( $x \geq 10$ ), and respectively regarding a time window of 30 minutes ranging from 49% ( $x \geq 2$ ) to 26% ( $x \geq 10$ ). If the minimization of the relative error is considered, each additional Bluetooth scanner decreases the error in average by 5% (10 or 30 minutes time window). If the time window duration is considered, the relative absolute error decreases by 9% in average while choosing a window size of 10 minutes respectively 30 minutes with the same number of Bluetooth scanners available. The error decreases is significantly more when  $x \geq 2$  (16%) as if  $x \geq 10$  (3%) is considered.



**Figure 5.7.** a) Estimation result with *basic* crowd density features. b) Estimation result with *advanced* crowd density features. a+b) Each ‘+’ symbol denotes a 10-minute snapshots from one of the 12 event areas. The x-axis shows the resulting linearly combined value of the feature vector, and the y-axis is defined by the ground truth value. The blue diagonal line denotes the 0% relative error value. The coloring visualizes the relative error of the predicted value (x-axis) regarding the actual value (y-axis). The range of the color scale is limited from 0.0 (0%) to 1.0 (100%) and values with a relative error larger than 1.0 are also colored red.

## 5.7 Crowd Motion Characteristics

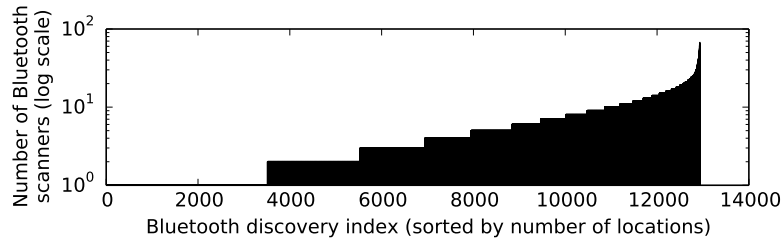
We evaluate the general feasibility to detect crowd flows in relying on mobile Bluetooth scanning devices and present qualitatively results by matching the extracted information to event schedule ground truth. The motivation of Bluetooth based crowd flow sensing is based on the assumption that just a few actively participating Bluetooth scanners are needed. Other surrounding people contribute indirectly

just with their enabled Bluetooth radio module without the need of an explicitly installed application. For example, if a single Bluetooth scanner senses five surrounding devices the statistical validity is higher than if a single sensor is measuring only its own movement.

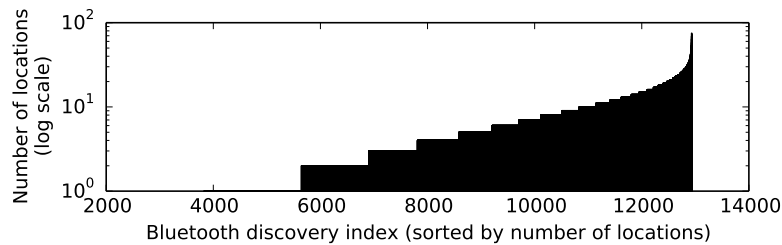
Bluetooth scanners might be static or moving. Those can be selected dynamically by their current association (which might change) to a certain area. The crowd flow can be measured between two areas. An area acts as a *virtual checkpoint* where Bluetooth devices are discovered. If a Bluetooth device identifier is discovered at another checkpoint a transition from area A to area B can be determined. Each transition detection is attributed with the duration of the transition. Virtual checkpoints can cover corridors (i.e. bridges, underpasses, streets) or any other places (larger and more distant areas) with numerous paths between two locations. While the latter might be interesting for analyzing patterns in visitor flows for marketing reasons, the former is most interesting for real-time analysis of emergencies in crowded areas. The tragic example of the Love Parade 2010 in Duisburg (Germany) demonstrated that such bottle necks can lead to fatal accidents.

We demonstrate the general feasibility to detect crowd flows in relying on mobile Bluetooth scanning devices by Figure 5.8 which visualizes all Bluetooth device identifiers compared to the number of Bluetooth scanners which discovered a Bluetooth identifier. In total 12933 discovery identifiers are involved. Around 3700 Bluetooth identifiers were discovered by just one scanner. In contrast some Bluetooth identifiers were discovered by up to 67 scanners. Figure 5.9 visualizes all Bluetooth device identifiers compared to the number of locations they were re-discovered. Locations are defined individually for a Bluetooth identifiers. A new location is represented by a discovery which is at least 10 meters away from any other discovery of the same Bluetooth identifier. Nearly 6000 Bluetooth discovery identifiers were not re-discovered at another location. In contrast around 7000 were re-discovered at another location at least once, with a maximum number of 75 locations.

We studied the crowd flow on the ‘Quai’ bridge which acted as a corridor between the western and eastern part of the city and as a spectator area at the same time. Figure 5.10 shows a satellite view with marked zones we used for transition monitoring. We aggregated all available scanners in each zone and time window.



**Figure 5.8.** Bar chart visualizing the Bluetooth discovery identifiers (x-axis, sum: 12933 discovery identifiers) which were discovered by different Bluetooth scanners (y-axis, log scale). The values are sorted by the ascending number of Bluetooth scanners. Wide bars are not to be confused with a single bar, but many contiguous equally high bars. 3700 Bluetooth discovery identifiers were discovered by just one scanner, while the remainder was discovered by at least two different Bluetooth scanners (maximum 67 Bluetooth scanners discovered the same Bluetooth discovery identifier).

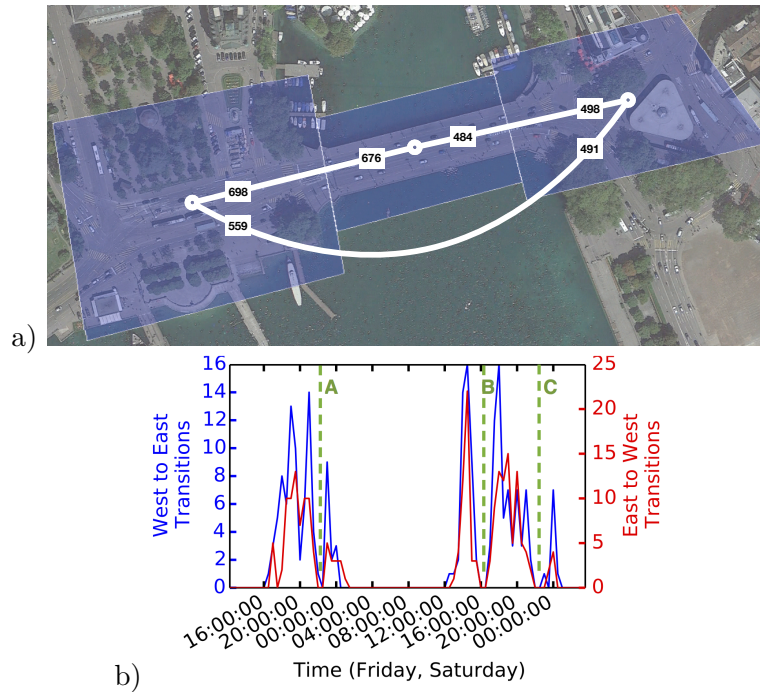


**Figure 5.9.** Bar chart visualizing the Bluetooth discovery identifiers (x-axis, sum: 12933 discovery identifiers) which reappeared (by proof of discovery) at least once and the respective number of different locations (y-axis, log scale) defined by a minimum distance of 10 meters to any previous discovery of the same identifier. The values are sorted by the ascending number of locations. Wide bars are not to be confused with a single bar, but many contiguous equally high bars. Nearly 6000 Bluetooth discovery identifiers were not re-discovered at another location, while around 7000 were re-discovered at least once (maximum location of re-discoveries: 75 locations).

The extracted transition information on the ‘Quai’ bridge during two days is shown in Figure 5.10 where separate time series show the crowd flow from the west to the east and the other way around. The time series simply represent the count of consecutive Bluetooth identifier observations within a 30 minute time window. Figure 5.10 qualitatively reveals a matching between the time series and the event

schedule ground truth. A strong connection between event times where spectators were stationary or walking around could be determined with the event calendar ground truth information. During three major events on lake Zurich the crowd flow on the monitored region quickly declined. At time (A) fireworks with music were presented (ground truth event calendar entry: Friday 10:30 p.m. to 11 p.m.), at time (B) a skydiver show was performed (ground truth event calendar entry: Saturday 4 p.m. to 4:30 p.m.), and at time (C) an acrobatic show was performed (ground truth event calendar entry: Saturday 10 p.m. to 10:15 p.m.).

This underlines our hypothesis that besides Bluetooth based crowd sensing we can also achieve Bluetooth based crowd flow monitoring with mobile Bluetooth scanners. This observation is new compared to previous literature where stationary Bluetooth scanners were used.



**Figure 5.10.** Indirect crowd flow measurement visualization. a) Diagram showing Bluetooth re-discoveries on the Quai Bridge. The numbers towards the end of a line section represent the number of Bluetooth devices re-appearing afterwards at the target zone. The blue colored regions represent the detection area. b) Time series of the crowd flow from the west (left blue area in Fig.a) to the east (right blue area in Fig.a). (A) fireworks (ground truth event calendar entry: Friday 22:30 to 23:00), (B) skydiver show (Saturday 16:00 to 16:30), and (C) acrobatic show. The maximum time lag between consecutive observations of the same device at different locations are 30 minutes. A strong peak can be detected on Friday during the arrival time between 6 p.m. and 9 p.m. and at around 11 p.m. during the departure time.

## 5.8 Citywide Grid-based Crowd Density Estimation

In the previous section we presented and validated the crowd condition estimation within pre-defined areas during the Zurich event. In this section we present a citywide grid-based crowd density estimation method and validation. However, the grid-based density estimation is not a real-time method for all density ranges (except identifying high crowd density levels) and focuses on a daily period crowd



density estimation.

### 5.8.1 General Considerations

A previously defined area has the size of thousand to tens of thousands of square meters. Areas are defined manually as being thematically associated. However, limitations exist when monitoring the crowd condition at semantically pre-defined areas. Areas have to be defined manually. An area is defined with the background knowledge. Coherency within an area must be given and known in advance of an event. Another disadvantage is that there is no guarantee of a uniform crowd distribution within an area - even when expected. Another disadvantage is that the crowd condition estimation within an area is aggregating all scanning data within the area. There is no knowledge of the exact location of the underlying scan coordinates after the aggregation.

In this section we propose the method of grid (or bin) based crowd condition estimation with Bluetooth scanning participators. Several advantages of grid-based crowd condition estimation opposed to area based crowd condition estimation exist. An advantage is the detailed resolution of crowd condition estimation entities. Diverging crowd conditions within an area can be estimated. No manual definition of event areas is necessary. Dynamically evolving event area boundaries are unproblematic. A disadvantage is the unsteady distribution of the crowd condition across neighbored entities when not enough scanning samples are available. An entity without scan samples is marked as void. The grid-based crowd condition estimation has to be seen as an evolving estimation method over time. The full picture of the crowd condition distribution is created over time. Participatory scanners must scan exactly within an entity to update its crowd condition. More frequented entities contain more scans and may contain newer scans. Less participatory scanner frequented entities contain less scans and contain potentially older scans. Grid-based crowd condition estimation is based on uniformly distributed entities along a lateral and longitudinal grid. Given a start entity, all other entities are deterministically distributed across the world. An entity can be defined as a polygon of common elementary categories such as a square, circle or hexagonal bin. We describe the attributes of the entities regarding its effectiveness for grid-based

crowd condition estimation. When using circles, grid aligned circles would leave holes in-between or otherwise would overlap when holes are avoided. Overlapping is critical because a non-ambiguous allocation to an entity is not possible. When using squares, grid aligned squares would perfectly align and would not leave holes and would not overlap. Lastly, when using hexagons, grid aligned hexagons would perfectly align and would not overlap. A hexagon grid consists of hexagons shifted by the half diameter in vertical direction. A hexagon based grid has advantages over a square based grid. Assuming a location in the center of the given entity, the euclidian distance to another location in the corner (north-east, south-east, south-west, north-west) is much further away as another location in the north, east, south or west. Another disadvantage of the square-grid is the problematic allocation of diagonal trajectories (north-east to south-west or south-east to north-west or inverse). A arbitrary diagonal participatory scanner walking pattern would yield to a staircase shaped allocation pattern to square entities. A hexagonal grid reduces the disadvantages of a square based grid. In a hexagonal grid the distance from the center to the boundary is nearly identical. Diagonal and straight movement trajectories are predominantly allocated to a straight sequence of hexagon entity. We defined the size of each hexagon entity with the background knowledge of the Bluetooth signal range. The specified hexagon diameter is reflecting the theoretical wireless Bluetooth signal range in a free line of sight condition. The Bluetooth range is limited to 10 meters. The hexagonal bin diameter is defined as 20 meters which is two times the Bluetooth range. Assuming a center location of a scanner within a hexagonal bin, the entire bin area is within the signal range. As the scanned signal source location is not known, the heuristically approach is expecting it and the signal attenuating crowd within the bin. We divided the whole central city of Zurich into 7838 hexagonal bins with a diameter of 20 meters.

In the following subsection we describe the crowd condition estimation methods and present evaluation results for the experiment during the citywide event in Zurich.

### 5.8.2 Definition of Crowd Class and Ground Truth

It is desirable to have a detailed ground truth information about the citywide crowd condition of the event. Detailed crowd condition (crowd density and crowd velocity) allows detailed temporal (i.e. at every second) and at the same time detailed spatial information (i.e. at every square meter). However, obtaining such detailed ground truth information for a whole city is infeasible. Similar to the participatory Bluetooth scanners we had other participators delivering their GPS position as a ground truth reference. GPS participators are independent of Bluetooth scan participators. GPS participators allowed continuous and anonymous tracking during the experiment within the city boundary. In total, we collected GPS traces from 29 800 people which were roaming throughout the city and reporting as ground truth indicator with their GPS location information over three days. We assume that the people are a representative sample. The fraction of GPS scanners is 1.3% according to the total number of 3.2 million event visitors. The fraction of participatory Bluetooth scanners is just 0.04%. 1.3% of the event visitors (every 77th person) are participatory GPS data collector. Likewise, at low frequented locations -during a whole day- no or only a very few intermittent participatory Bluetooth scanners might appear. To maximize the range of crowd density classes available for the method validation we selected a whole day for the aggregation time window. Within the duration of one day we collected a detailed spatial resolution of the crowd condition. With this method we rely on GPS information of up to 3000 people per hexagonal bin. We define the ground truth ‘crowd level’ as the connection between the quantity and the speed of the GPS participators. The quantity per bin is defined as the number of distinct GPS participators localized in hexagonal bin ( $\xi_x$ ). This quantity alone does not represent the crowd condition very well over the duration of one day. Bins through which people roam over time are marked as a high quantity (see Figure 5.11b on page 170). The most frequently used pathways throughout the city are recognized by this value. Bins with high crowd density and many transitions are not recognizable. To compensate this effect we aggregate the crowd speed in each hexagonal bin ( $\nu_x$ ). We only rely on GPS participators with a maximum velocity of 2.5 m/s which is equal to running speed. Faster moving transportation methods are not included. The crowd speed is

computed by averaging all filtered GPS participants velocities while being observed in the hexagonal bin during the day. Figure 5.11c is visualizing the crowd speed in each hexagonal bin. We see that the city center is dominated by a low crowd speed while the adjacency bins are dominated by crowd speeds above 0.7 m/s. The Zurich lake is a special case where slowly moving ships are underway with a low velocity of around 1.6 m/s. Neither the quantity nor the velocity are reflecting the crowd condition very well on their own. But the combination gives an important statement.

Low crowd velocity and high quantity would indicate a highly crowded bin. High crowd velocity and high quantity would indicate a medium crowded bin. A high crowd velocity and low quantity would indicate a low-crowded bin. To emphasize bins which are not often visited by crowd scanners we apply logarithm function. To emphasize bins with just few scanners but with low crowd speed -assuming the low speed is based on a high crowd density- we apply the exponential function. We define the ‘crowd level’ by the function  $f$ :

$$f(\text{hexagonal bin}_x) = \log(\xi_x) / \exp(\nu_x)$$

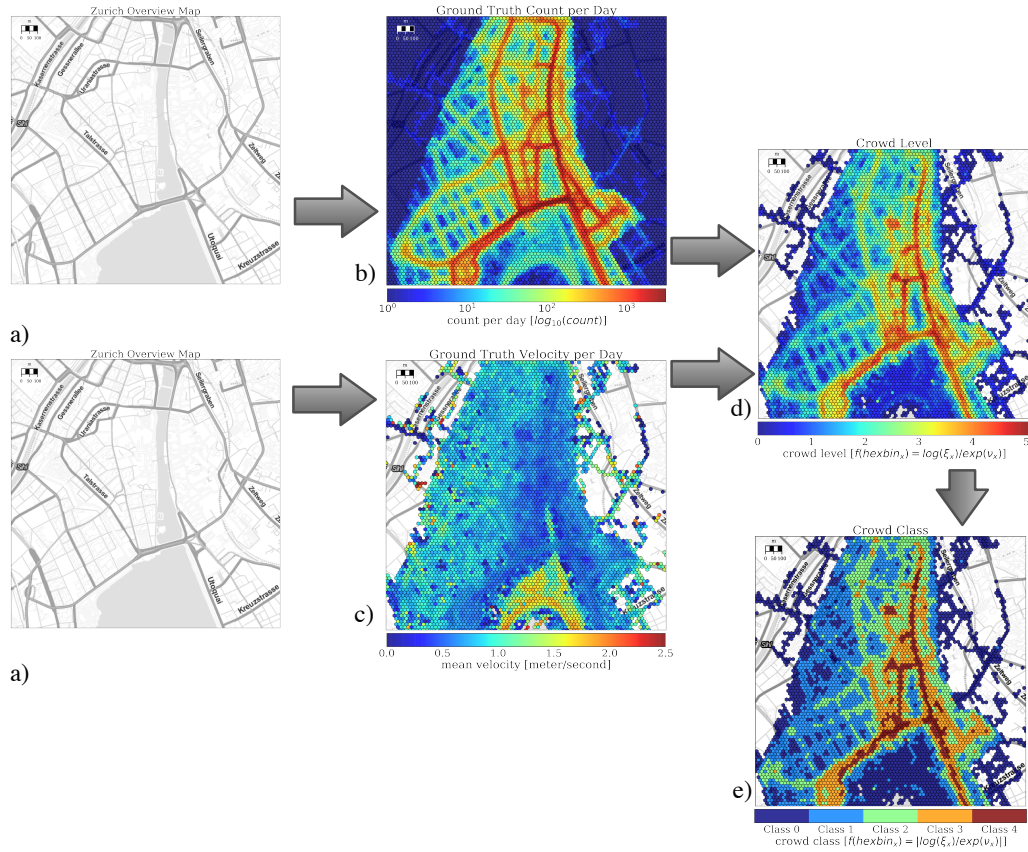
Figure 5.11d is visualizing the crowd level during one day of the Zurich event. To prepare the crowd level for machine learning classification we discretize the crowd level values into five discrete ‘crowd classes’ ranging from discrete class 0 (uncrowded) to class 4 (crowded). The ‘crowd classes’ are visualized in Figure 5.11e.

### 5.8.3 Methods and Machine Learning

We propose methods to estimate the crowd class within a grid of 7838 hexagonal bins with a diameter of 20 meters each. The methods build on the foundation of machine learning classification. Analogous to the time frame based ground truth extraction the methods estimate the ‘crowd level’ for a full event day. To allow comparability with the ground truth we compute the following features for the same time window.

The first proposed method is based on quantities of values being observed (absolute features) similar to  $t$ . The second method is based on measurable variations of

## 5 Participatory Citywide Sensing



**Figure 5.11.** Definition of the ground truth ‘crowd level’. Defined by the quantity of GPS sensor participators and the velocity of the GPS sensor participators per hexagonal bin. b) The quantity ( $\xi_x$ ) per bin. c) The crowd velocity ( $\nu_x$ ) per hexagonal bin. d) The combined crowd level by the formula  $f(\text{hexagonal bin}_x) = \log(\xi_x)/\exp(\nu_x)$ . e) Discretized ‘crowd level’.

values being observed (relative features). As described in the previous chapters, absolute quantity methods are strongly influenced by the background of the surrounding people. Within one group of people everyone might be equipped with a wireless enabled device. In another group of people nobody might be equipped with a wireless enabled device. Such fluctuations are impossible to be detected for absolute features. We create a multi-dimensional feature vector to reduce such influences over time and when multiple scanners were active in a hexagonal bin. Relative features on the other hand measure variations which are produced by

signal attenuations by the people themselves and not due to the background of the people. We introduce multiple independent relative features combined to a multi-dimensional feature vector.

Each feature is computed for each of the 7838 hexagonal bins. To identify the currently associated bin of each scanner we rely on location information from the participatory scanner (consecutively called scanner). The scanners sent location information at varying time intervals. When being static, location information was sent every 30 seconds. When moving, location information was sent every few seconds. Situations in which GPS beams from the satellites were shielded did not allow location extraction. Such situations occur under bridges or in narrow streets between tall houses. When location information was not available during the Bluetooth scan the Bluetooth scan was ignored. Bluetooth scans were performed once every minute. A Bluetooth scan includes a set of distinct Bluetooth devices. A Bluetooth scan does not allow insight into the continuity of the devices within the scan. No assertion is possible whether a device is discovered while passing by or whether the signal attenuation is varying for a single device. Due to the growing size of the data set over the experiment day, one can assess different devices' signals scanned by multiple scanners over time within each bin.

The absolute features are defined as the 1) total number of unique detections per hexagonal bin, 2) the total number of detections including repeated detected devices per hexagonal bin, 3) the average number of detected devices per scanner per hexagonal bin and 4) the number of scanners being active in the hexagonal bin. The feature of the total number of unique detections per hexagonal bin is created by the size of the set of distinct discovered devices over time. This feature is independent of the repeatedly discovered devices but models the quantity of distinct detections and the diversity of detections. Over time the set size is increased as new distinct devices are discovered by the same or by other scanners within the bin. A single well receivable device's wireless signal won't influence this feature when multiple scanners detect the same device multiple times. In our mobile scanning scenario scanners and discovered devices are mobile and are potentially roaming throughout the commemorated bin boundaries. When the discoverable device is roaming to another bin and is re-detected it is again considered in the set of unique devices per bin. The feature of the total number of detections including repeated

detected devices per hexagonal bin is created by the size of the array of discovered devices over time. This feature includes potential multiple detection of the same device being stationary in the bin. This feature describes the number of scans within a bin even when both scanner and discovered device are not moving the value is increased. This feature is important for bins with less pedestrian flow, i.e. during a concert when people are standing still in a dense crowd and no new unique devices are arriving in the allocated bin over the duration of the concert. The feature of the average number of detected devices per scanner per hexagonal bin is created by the size of detected devices per scanner within the bin divided by the number of scanners being active within the bin. This feature describes the average detections of devices per scanner within one scan. This feature compensates divergences between multiple scanners and over time. With an increasing duration of observed scans this feature is representing the expected number of devices being discovered during one scan window within the bin. The feature of the unique number of scanners in the hexagonal bin is created by the set size of the number of unique scanners being active at least once within the bin. This feature describes the visit frequency of scanners in the bin. This feature is independent of the duration of the scanner visit within the bin and is independent of the number of scanners re-visiting a bin again for a second time or more.

The relative features are defined as 5) the signal variation over all devices being detected over time including re-discoveries in the bin, 6) the mean signal strength over all devices being detected over time including re-discoveries in the bin, 7) the variation over the quantity of devices per scan along all scanners within the bin, 8) the mean scanner velocity during the scans in the bin, and 9) the fraction of scans returning no discovered device in the bin against the total number of scans performed in the bin.

The feature of the signal variation over all devices being detected over time in the bin is created by the variance between the measured received signal strength indicators of all detected devices including re-discoveries of the same device during another scan by the same scanner or by any other scanner. The variance is defined as  $variance_{signal}(bin) = \frac{\sum_{x \in bin} signal^2}{number_{bin}} - average(bin)^2$  where  $x$  describes a device discovery with a connected signal strength measurement. This feature takes signal strength variations into account which are induced by varying signal attenuations

at different crowd densities and varying distances between scanner and discovered devices. While the true distance to a device is unknown we assume that the distances are uniformly distributed as time continuous and further scanning devices are discovered. We expect stronger variations in crowded environments than in non-crowded environments. The feature of the mean signal strength of all detected devices including re-discoveries of the same device during another scan by the same scanner or by any other scanner is defined as  $mean_{signal}(bin) = \frac{\sum_{x \in bin} signal}{number_{bin}}$ . This feature takes the signal strength into account which is induced by signal attenuation at different crowd densities and the distance between scanner and discovered device. Over time we expect a lower signal strength in crowded environments than in non-crowded environments. The feature of the variation over the quantity of devices per scan in the bin is created by the variance of the number of detected devices per scan by the same scanner or any other scanner. The variance is defined as  $variance_{quantity}(bin) = \frac{\sum_{x \in bin} quantity^2}{number_{bin}} - average(bin)^2$  where  $x$  describes a the number of unique devices within a scan. This feature takes variations of the number of devices detected per scan into account. We expect a stronger variation in lower crowd densities than in higher crowd densities. In low crowd densities it is expected that it strongly depends on the group of people currently being in the proximity of a scanner. Within one group every person might be equipped with a wireless Bluetooth enabled device, within another group no person might be equipped with a Bluetooth device. When the number of persons in the proximity is high we assume that different groups are within the range and the number of Bluetooth enabled devices is uniformly distributed. The feature of the mean scanner velocity during the scans in the bin is created by the sum of the velocity during a scan divided by the number of scans performed. This feature takes the potential ability of detecting new unique devices into account. A scanner can either detect many devices when moving fast through a less dense crowd or can detect many devices when walking slowly through a dense crowd. This feature compensates the effect of the scanner velocity while the velocity of discovered devices in the proximity is unknown. The feature of the fraction of scans returning no discovered device in the bin against the total number of scans performed in the bin is created by the number of scan intervals without detecting any device divided by the number of performed scans with zero or more detected devices in the bin. This feature takes



the scanning into account that tried scanning for devices in the bin but did not resolve any devices. This feature emphasized the detection of at least one device against the detection of none device at all. It is expected that bins with very low crowd densities have many empty scans over time. Because of the assumed uniform distribution of scanners within the city crowded bins are covered more often than non-crowded bins. This allows to support the assumption of bins being not crowded at all.

Our methods to estimate the ‘crowd level’ rely on machine learning of the data records  $(x, Y)$ , where  $x$  is the feature vector and  $Y$  is the target value. The general machine learning principle was described in detail in subsection 4.6.3. For machine learning we selected the decision tree classifier and configured its parameters to generally avoid overfitting. The decision tree classifier is configured to have a maximum depth of 15 from the root not to any leaf node and the decision tree is configured to have a minimum number of 15 data records per leaf. This configuration is common in decision tree learning where a large number of data records is available. The machine learning approach follows an independent training with a subset of the data while another subset of the data is used for validation. The classifier results are evaluated by 10-fold cross-validation. With this approach every data record is used once in the training set. The classification performance metrics are averaged along the cross-validation results.

In general, the classifier is trained by either the feature set of absolute features or by the feature set of relative features. One feature vector was computed for each of the 7838 hexagonal bins. Together with the ground truth ‘crowd level’ labels  $Y$  for each bin (as defined in the previous subsection) the data record  $(x_1 \dots x_k, Y)$  is created for each bin. During one cross-validation step one hexagonal bin data record either served as training or as validation instance. This means we divide the whole city geographically into training bins and validation bins. While iterating over the cross-validation steps each bin is used once for training. While a bin is used for training it is not used for validation as of the principle of machine learning. By having equally sized hexagonal bins as data records we avoid specializing on specific manually defined regions of the event, but generalize the classifier to be universally applicable to a dynamically extendable region.

For the evaluation of the influence of the quantity of participative scanners within

a bin, we selected a subset of currently considered scanners in a bin. For each subset of currently considered scanners the features are individually generated. For each bin we randomly selected  $i_x$  ( $i \leq n_x$ ) distinct scanners ( $n_x$  is the maximum number of scanners available in bin  $x$ ). We considered a range between a minimum of 1 scanner and maximum of 25 scanners per bin. This yielded into 24 individual feature computation runs. In total we have  $7838 \cdot 24 \cdot 9 = 1\,693\,008$  feature computation runs over all bins and subsets. Feature computation was done with parallel processing to reduce run-time. The absolute feature vectors and the relative feature vector are computed individually. According to the general classifier training procedure with cross-validation each subset is evaluated individually. In total we have  $2 \cdot 24 \cdot 10 = 480$  classifier runs.

#### 5.8.4 Evaluation Results

We evaluated our ‘crowd level’ estimation method for four scenarios. The first two scenarios allow insight into the estimation performance as achieved with the exact amount of all 981 scanners and their scanning data available during the Zurich event. The third and fourth scenario allow insight into the estimation performance when only a lower number of scanners is available. This is an important aspect when applying the estimation method to other events and to know the number of scanners needed within a bin for a certain performance metric. The scenarios include the validation of the

- absolute feature vectors from the full spectrum of data available by all scanners being active during the citywide event (scenario 1)
- relative feature vectors from the full spectrum of data available by all scanners being active during the citywide event (scenario 2)
- absolute feature vectors from several subsets of the data by artificially limiting the number of scanners being active in each hexagonal bin (scenario 3)
- relative feature vectors from several subsets of the data by artificially limiting the number of scanners being active in each hexagonal bin (scenario 4).

In Figure 5.12 the performance metrics are shown for a) the absolute feature vector and for b) the relative feature vector. The shown performance metrics are the weighted classification precision for each class, the weighted classification recall for

each class and the weighted classification f-score for each class. We use weighted precision and weighted recall to compensate the different number of data records per class. Intuitively a high precision is the ability of the classifier to return accurate results. Intuitively recall is the ability of the classifier not to label a sample as positive that is negative. The f-score is the harmonic mean of the precision and recall, allowing a one-value classifier performance description. However, we present the f-score as an add-on and use the precision and recall as the main classifier performance metric values. We present the evaluation results separately based on the absolute method and the relative method in the order of the aforementioned scenarios, continued by presenting the similarities and differences between the methods.

We begin with the scenario ‘1’ -the full spectrum of data available by all scanners- and the *absolute* feature vector. The performance metric results for the full data set are shown for the x-axis value of 24 in Figure 5.12a in each of the three sub-plots. The scenario ‘1’ has classification precisions (in increasing order of the precision) of 0.52 for class 1, 0.61 for class 2, 0.68 for class 0, 0.72 for class 3 and 0.80 for class 4. We observe that the classes 4, 3 and 0 have the best classification precision. These are the classes with either very high or very low crowd density. The most accurately estimated data records belong to class 4, class 3, and class 0. The other classes 1, 2 and 0 are not estimated accurately. We see in the later evaluation the confusion of the estimation. In average we have a classification precision of 0.64. The classification recall values are (in increasing order of the recall) 0.06 for class 1, 0.40 for class 2, 0.62 for class 3, 0.71 for class 4 and 0.99 for class 0. We observe that class 0 and the class 4 have a large fraction of true positives in relation to the sum of true positives and false negatives. This means that the classifier’s ability is proportionally high to not estimate the other classes as class 0 or class 4. This also means that the classes 1, 2 and 3 have a low fraction of true positives in relation to the sum of true positives and false negatives. In average we have a classification recall of 0.67.

We continue with the scenario ‘2’ -the full spectrum of data available by all scanners- and the *relative* feature vector. The performance metric results for the full data set are shown for the x-axis value of 24 in Figure 5.12b in each of the three sub-plots. The scenario ‘2’ has classification precisions (in increasing order of the precision)

of 0.54 for class 1, 0.64 for class 2, 0.67 for class 3, 0.68 for class 0 and 0.75 for class 4. We observe that the classes 4 and 0 have the best classification precision. This means that the class 4 and class 0 are often correctly estimated and at the same time a low number of data records is falsely classified as class 4 or class 0. In average we have classification precision of 0.65. The classification recall values are (in increasing order of the recall) 0.09 for class 1, 0.38 for class 2, 0.57 for class 4, 0.62 for class 3 and 0.99 for class 0. We observe that class 0 and the class 3 have a large fraction of true positives in relation to the sum of true positives and false negatives. This means that the classifier's ability is proportionally high to not estimate the other classes as class 0 or class 3. This also means that the class 1 and class 2 have a low fraction of true positives in relation to the sum of true positives and false negatives. In average we have a classification recall of 0.67.

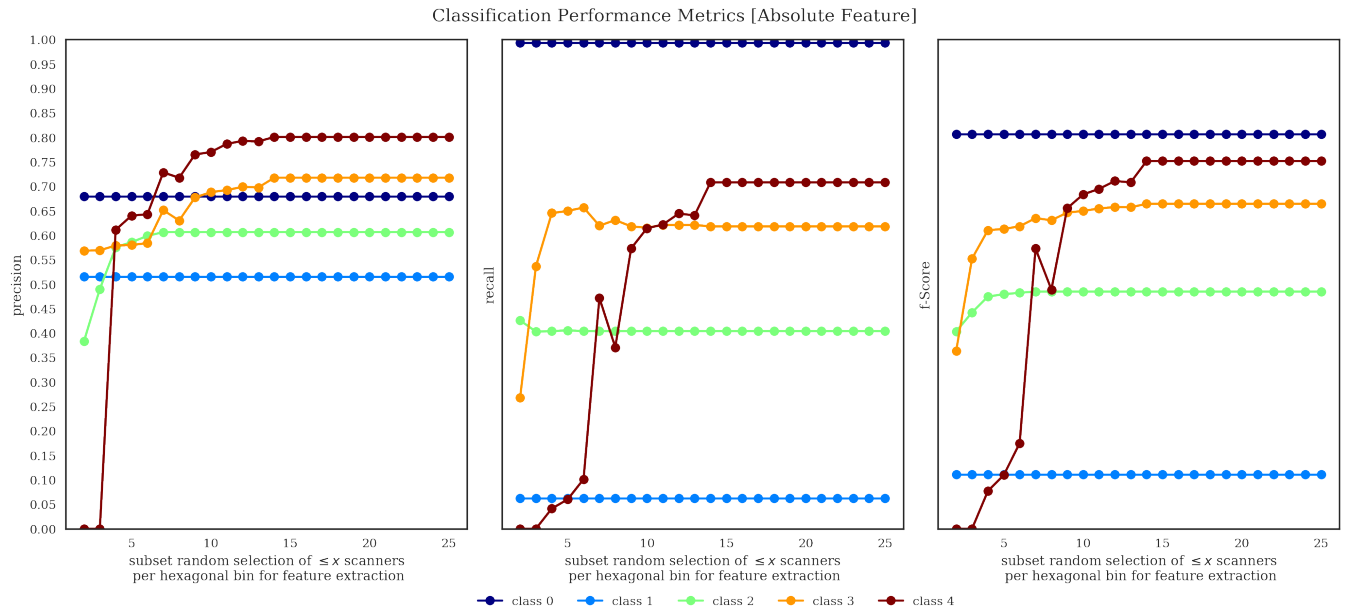
We continue with the scenario '3' with several subsets of the data by artificially limiting the number of scanners being active in each hexagonal bin and the *absolute* feature vector. The performance metric results for each of the subsets are shown for the x-axis values ranging from 2 to 25 in Figure 5.12a in each of the three sub-plots. By increasing the subset of active scanners  $s$  to  $s \leq x$  in scenario '3' we observe an increasing classification precision for the classes 4, 3 and 2. We observe that the classification precision is initially increasing strongly with a following convergence to the best possible precision. We see that for class 3 and 4 the optimum precision is nearly reached at  $s \leq 10$ . The optimum precision for class 2 is reached earlier at  $s \leq 7$ . The other classes 0 and 1 have a steady classification precision and increasing  $x$  does not change the classification precision. The classification recall is increasing for an increased value of  $x$  for the classes 3 and 4. We observe that the classification recall is initially increasing strongly with a following convergence to the best possible recall. We see that for class 3 the optimum recall is nearly reached at  $s \leq 4$ . The optimum recall for class 4 is reached at  $s \leq 14$ . For class 4  $s \leq 8$  falls below  $s \leq 7$ , while  $s \leq 9$  continues in the increasing trend. The other classes 0, 1 and 2 have a steady classification recall and increasing  $x$  does not change the classification precision.

We continue with the scenario '4' with several subsets of the data by artificially limiting the number of scanners being active in each hexagonal bin and the *relative* feature vector. The performance metric results for each of the subsets are shown for

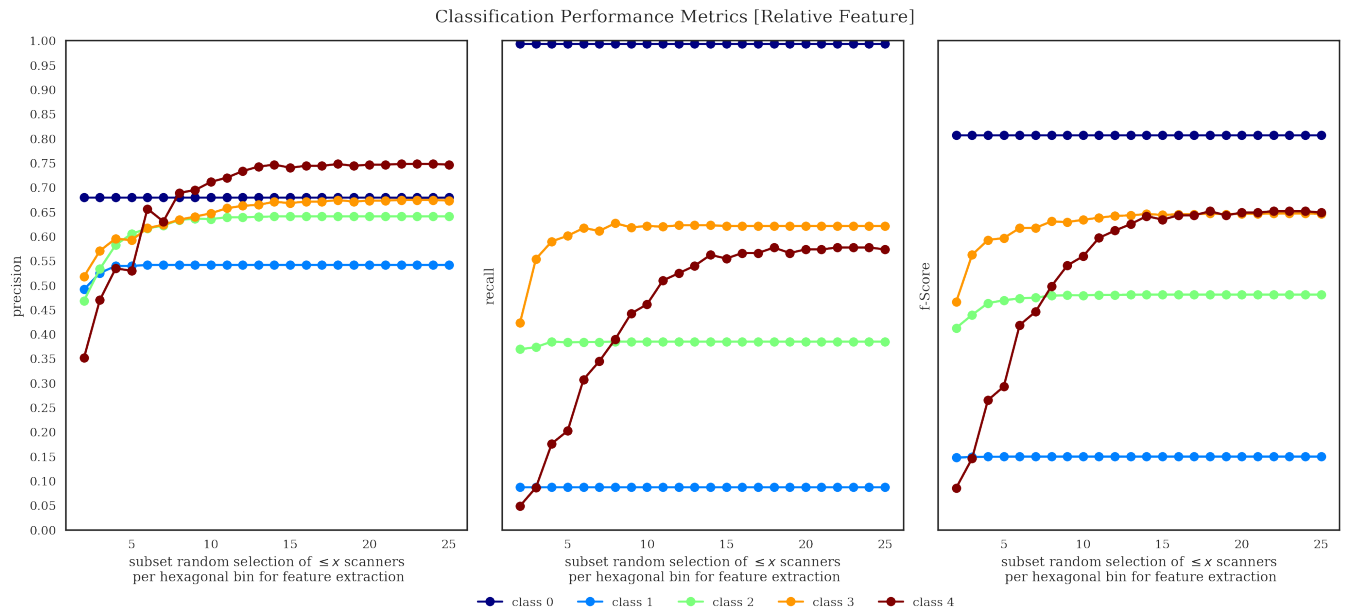
the x-axis values ranging from 2 to 25 in Figure 5.12b in each of the three sub-plots. By increasing the subset of active scanners  $s$  to  $s \leq x$  in scenario '4' we observe an increasing classification precision for the classes 4, 3, 2 and 1. We observe that the classification precision is initially increasing strongly with a following convergence to the best possible precision. We see that for class 1 the optimum precision is nearly reached at  $s \leq 4$ . The optimum precision for class 2 is reached nearly at  $s \leq 7$ . The optimum precision for class 3 and 4 is reached nearly at  $s \leq 12$ . The classification recall is increasing for an increased value of  $x$  for the classes 3 and 4. We observe that the classification recall is initially increasing strongly with a following convergence to the best possible recall. We see that for class 3 the optimum recall is nearly reached at  $s \leq 6$ . The optimum recall for class 4 is reached at  $s \leq 14$ . The other classes 0, 1 and 2 have a steady classification recall and increasing  $x$  does not change the classification precision.

We observe that the scenario '2' classification precision with the *relative* feature vector equals the precision of scenario '1' - and this even by not relying on the absolute count of detected Bluetooth devices. Comparing the scenario '3' and scenario '4' the classification precision has a similar progression over the increasing number of active Bluetooth scanners per bin.

## 5.8 Citywide Grid-based Crowd Density Estimation

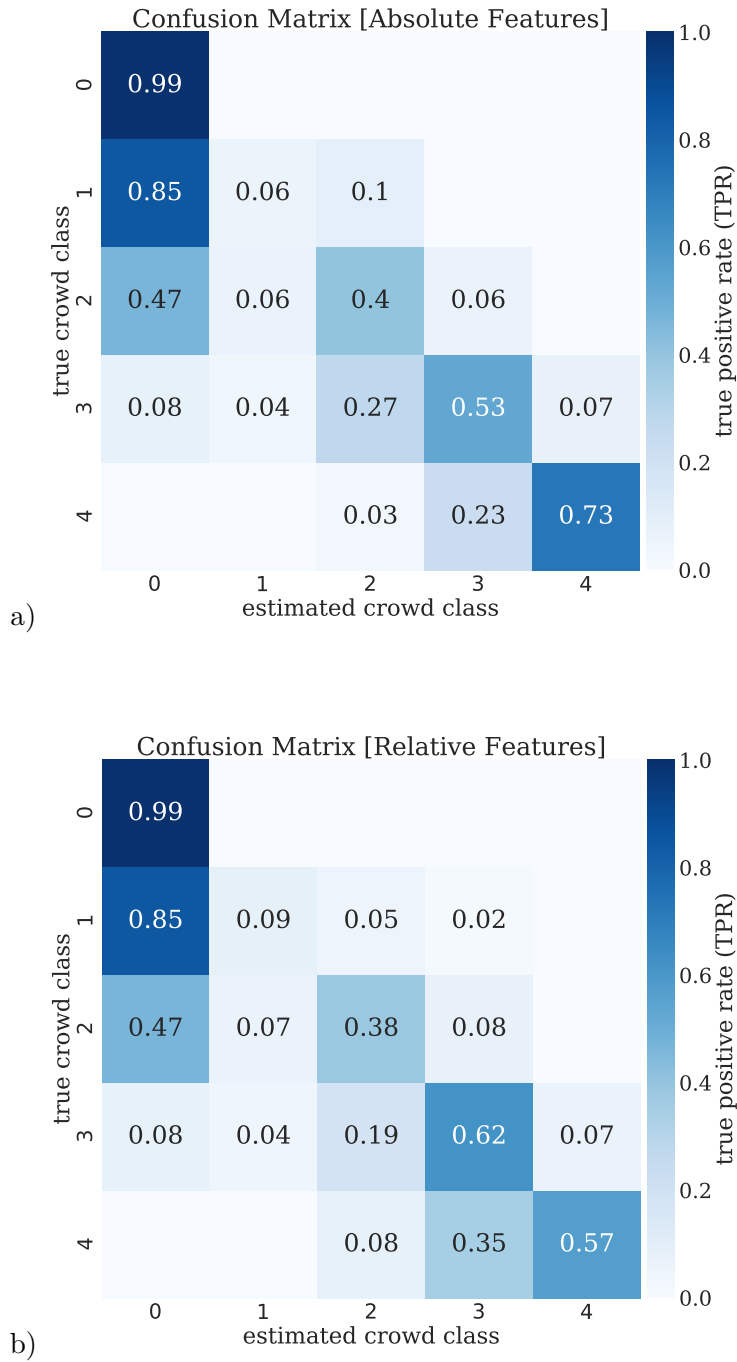


a)



b)

**Figure 5.12.** Classification performance metrics. Grid-based crowd level estimation evaluation results based on a) absolute feature vectors from several subsets of the data by artificially limiting the number of scanners being active in each hexagonal bin and the full spectrum of data available by all scanners. b) Relative feature vectors from several subsets of the data by artificially limiting the number of scanners being active in each hexagonal bin and full spectrum of data available by all scanners being active during the citywide event.



**Figure 5.13.** Confusion matrices. Grid-based crowd level estimation evaluation results based on a) absolute feature vectors on all participatory scanners, b) on relative feature vectors on all participatory scanners being active during the citywide event.

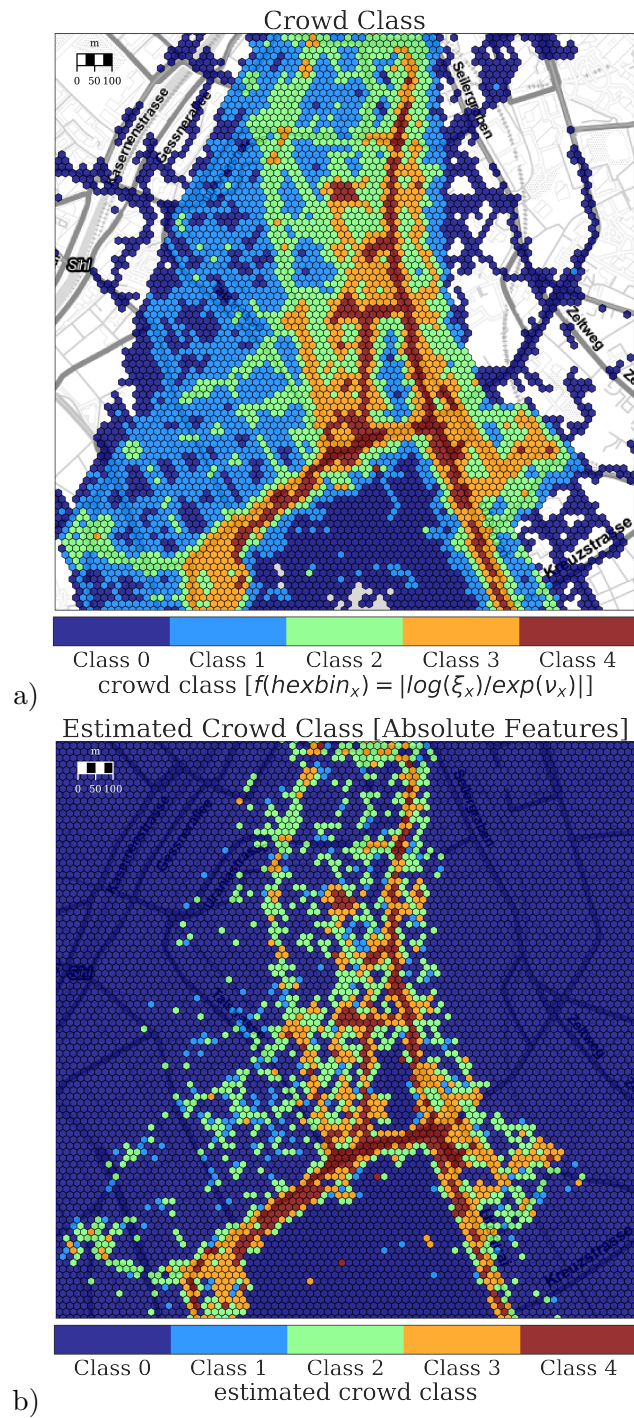
## Spatial Grid-based Results

In the following we present the results of the spatial distribution of the estimation results of the hexagonal bins. The question is whether we can observe the spatially structure of the ground truth crowd class again in the estimated crowd class.

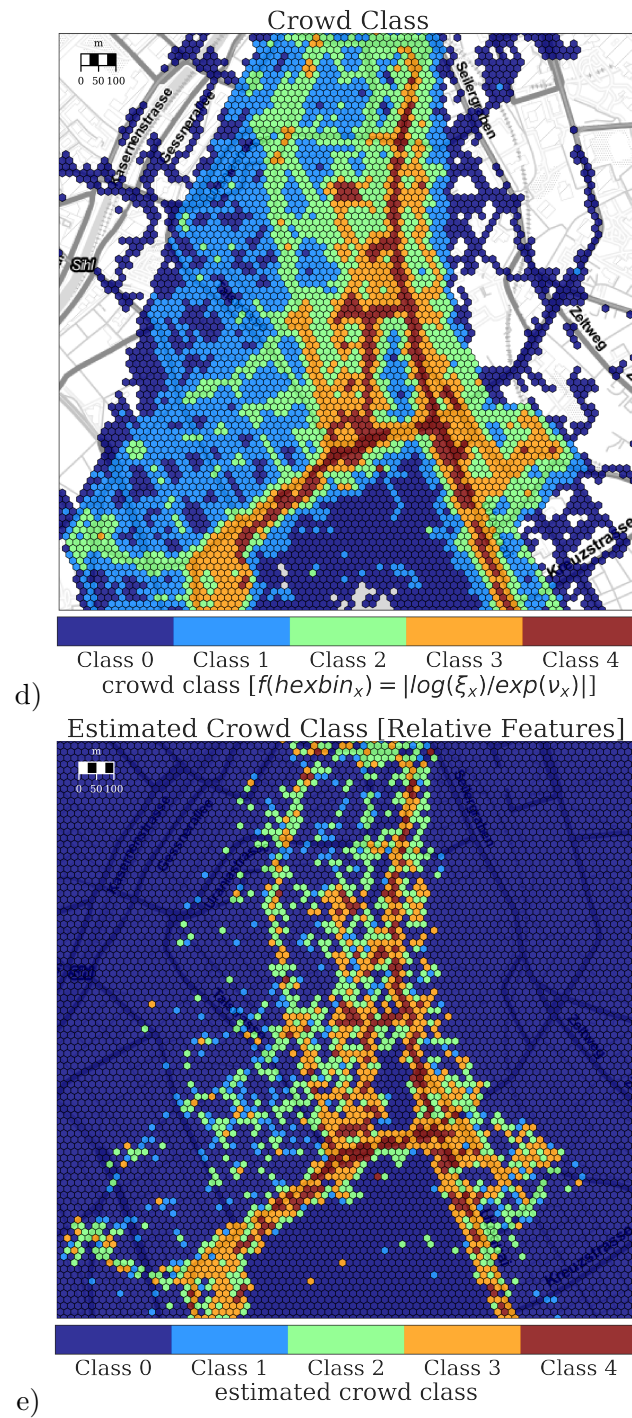
In Figure 5.14 we visualized the scenario ‘1’ spatial distribution of the ground truth crowd classes and the spatial distribution of the estimated crowd classes. We observe that the hexagonal bins of crowd classes 1, 2 and 3 are dominantly correctly estimated and mis-classified as class 0. The crowd classes 3 and 4 are well estimated and represent the same spatial structure as the ground truth distribution. Some hexagonal bins of class 4 are just classified as the very similar class 3. We can recover the spatial structure of the ground truth classes 3 and 4 very well. The confusion matrix in Figure 5.13a is confirming this observation. We see that the true crowd class 1 is confused by 85% to crowd class 0 and by 10% to crowd class 2.

In Figure 5.15a we visualized the scenario ‘2’ spatial distribution of the ground truth crowd class and the spatial distribution of the estimated crowd classes. We observe that the hexagonal bins of crowd classes 1, 2 and 3 are again intermittent correctly estimated and often classified as class 0. The crowd classes 3 and 4 are well estimated. Some hexagonal bins of class 4 are just classified as the very similar class 3. We can recover the spatial structure of the ground truth classes 3 and 4 very well. The confusion matrix in Figure 5.13b is confirming this observation. We see again the confusion of the true crowd classes 1, 2 are confused by 85% to crowd class 1 or 47% to crowd class 2. The key difference in the confusions between scenario ‘1’ and scenario ‘2’ are the mis-classifications of the classes 3 and 4. Scenario ‘1’ has less confusion crowd class 4 than scenario ‘2’. Scenario ‘2’ has less confusion crowd class 3 than scenario 1. For both methods the most confusion according to Figure 5.13a and Figure 5.13b are with the classes 1 and 2. Class 1 is mostly confused with the neighbored class 0. Class 2 is confused often with the class 0. We suspect that the medium crowded classes 1 and 2 are prone to the presence or absence of wireless enabled groups of people. The classes 0, 3 and 4 are the least confused.





**Figure 5.14.** Spatial visualization of crowd level estimation in hexagonal bin scale. Absolute feature vectors on all participatory scanners.



**Figure 5.15.** Spatial visualization of crowd level estimation in hexagonal bin scale. Relative feature vectors on all participatory scanners.

## 5.9 Conclusion

The work presented in this chapter clearly shows the potential of using Bluetooth scanning as means of monitoring crowds in urban environment. The accuracy of our density estimation is well within needs of typical crowd management applications. The analysis of the data set (section 5.5) has demonstrated that the analysis strongly depends on a relatively small number of highly mobile nodes. This means that, for more spatially constrained events just equipping the security personnel with scanners may be enough. Similarly, the technique could support urban crowd monitoring outside specific events (as demonstrated in section 5.8). To this end one would need to recruit volunteers who regularly move around the city a lot. How many people need to be recruited and what mobility patterns would be needed is something that would have to be studied empirically in a real life experiment. Building on insights from section 5.5, section 5.6, and section 5.8 future work could investigate which types of users contribute to the result better, whether different patterns can be extracted at different time periods, and to what degree aspects such as the distribution of the number of found devices per scan and the temporal and spatial distribution of the reoccurrence of devices in the scans can be used to reason about the character of events on a more complex level.

# List of Figures

1.1	Ontology of wireless signal based crowd condition estimation methods	8
2.1	Thesis ontology - chapter 2	14
2.2	Illustration of meshed stationary sensors passively scanning crowd centric devices	14
2.3	Bluetooth inquiry response packet payload structure	23
2.4	WiFi probe packet payload structure	25
2.5	Photographs of scanner unit and setup preparations	34
2.6	Stationary scanner setup during the exhibition IAA (Frankfurt, Germany)	35
2.7	Regression results in comparison with ground truth (IAA experiment)	49
2.8	Time series of the estimated crowd density (IAA experiment)	50
2.9	Heat-map of the crowd density (IAA experiment)	56
2.10	Heat-map of the crowd density (IAA experiment)	57
2.11	Heat-map of the crowd flow (IAA experiment)	58
2.12	Contextual anomaly detection press conference (IAA experiment)	62
2.13	Contextual anomaly detection chancellor Dr. Merkel visit (IAA experiment)	63
3.1	Thesis ontology - chapter 3	65
3.2	Symbolic illustration of opportunistically scanning ambient stationary devices	70
3.3	Scanner setup and ground truth camera setup (CeBIT experiment)	74
3.4	Data transformation process and time series results of static ambient device signal scanning (CeBIT experiment)	75

*List of Figures*

3.5	Multi-path and blocking line of sight . . . . .	77
3.6	Sensing unit for special purpose connection-oriented CSI scanning	78
3.7	CSI static people counting - fingerprinting complexity evaluation (DFKI experiment) . . . . .	85
4.1	Thesis ontology - chapter 4 . . . . .	90
4.2	Mobile scanning of opportunistic crowd-centric devices (Oktoberfest experiment) . . . . .	91
4.3	Histogram with distribution of mobile scans across venues . . . . .	97
4.4	Ground truth crowd density reference (Oktoberfest experiment) . .	107
4.5	Data distribution of unique Bluetooth detections with individual and group scanning . . . . .	108
4.6	Confusion matrices mobile individual scanning (Oktoberfest experi- ment) . . . . .	110
4.7	Confusion matrices mobile group scanning (Oktoberfest experiment)	112
4.8	Link structure between mobile scanners and discovered devices (Kaiserslautern public viewing experiment) . . . . .	113
4.9	Course of the crowd density (Kaiserslautern public viewing experiment)	114
4.10	Course of the crowd flow (Kaiserslautern public viewing experiment)	115
4.11	Crowd density classes (Kaiserslautern public viewing experiment) .	122
4.12	Feature ‘Size of device set of all distinct discovered devices by all sensors in time frame’ (Kaiserslautern public viewing experiment) .	123
4.13	Feature ‘Ratio bi-directional link structure of sensors to average pairwise distance of sensors multiplied with average sensor speed’ (Kaiserslautern public viewing experiment) . . . . .	124
4.14	Feature ‘Ratio of discovered devices in current snapshot to discovered devices in last x minutes’(Kaiserslautern public viewing experiment)	124
4.15	Feature ‘Average team-wise diversity of discovered devices per scan window (ratio not concurrent to concurrent)’ (Kaiserslautern public viewing experiment) . . . . .	125
4.16	Feature ‘Average number of semi-continuous unique device visibility periods’ (Kaiserslautern public viewing experiment) . . . . .	125

4.17	Feature ‘Average durations of semi-continuous unique device visibility periods (finite state machine approach)’ (Kaiserslautern public viewing experiment) . . . . .	126
4.18	Confusion matrix collaborative absolute features results (Kaiserslautern public viewing experiment) . . . . .	130
4.19	Confusion matrix collaborative relative features results (Kaiserslautern public viewing experiment) . . . . .	131
4.20	Confusion matrix collaborative absolute and relative features results (Kaiserslautern public viewing experiment) . . . . .	132
5.1	Thesis ontology - chapter 5 . . . . .	135
5.2	Experiment area definition and ground truth heat-map snapshot (Zurich participatory experiment) . . . . .	137
5.3	Impressions during the event (Zurich participatory experiment) . . . . .	138
5.4	Bluetooth discovery structure graph topology (Zurich participatory experiment) . . . . .	148
5.5	Distribution of participatory Bluetooth scanners (Zurich participatory experiment) . . . . .	149
5.6	Histograms of the Bluetooth scan distribution during multiple experiments . . . . .	153
5.7	Estimation result distribution (Zurich participatory experiment) . . . . .	161
5.8	Bar chart visualizing the ratio between specific discovered devices and the quantity of scanners (Zurich participatory experiment) . . . . .	163
5.9	Bar chart visualizing the re-appearance rate of specific devices at other locations (Zurich participatory experiment) . . . . .	163
5.10	Indirect crowd flow estimation (Zurich participatory experiment) . . . . .	165
5.11	Grid-based ground truth in citywide sensing (Zurich participatory experiment) . . . . .	170
5.12	Estimation performance metrics for grid-based citywide crowd condition estimation (Zurich participatory experiment) . . . . .	179
5.13	Confusion matrices, citywide grid-based estimation (Zurich participatory experiment) . . . . .	180
		187

*List of Figures*

5.14 Hexagonal bins with absolute feature results (Zurich participatory experiment) . . . . .	182
5.15 Hexagonal bins with relative features results (Zurich participatory experiment) . . . . .	183

# List of Tables

1	Performed crowd condition estimation experiments. . . . .	
1.1	Selected list of publications . . . . .	12
2.1	Experiment data sets statistics of stationary opportunistic crowd device scanning . . . . .	30
2.2	Scanned WiFi and Bluetooth crowd device vendors statistics . . . . .	31
2.3	Machine learning crowd density estimation errors . . . . .	47
2.4	Resulting coefficients for local Voronoi cell based crowd density estimation . . . . .	52
3.1	Evaluation results people count large scale ambient signal variations	73
3.2	CSI and RSSI static people counting results (DFKI experiment) . . . . .	82
3.3	CSI static people counting - comparison statistical reference method (DFKI experiment) . . . . .	85
3.4	CSI static people counting - antenna setup evaluation (DFKI experiment) . . . . .	86
3.5	CSI static people counting - feature evaluation (DFKI experiment)	87
3.6	CSI static people counting - people range evaluation (DFKI experiment) . . . . .	87
4.1	Related work crowd condition estimation concepts (Oktoberfest experiment) . . . . .	95
4.2	Data set statistics mobile scanners . . . . .	98
4.3	Discovered mobile device categories and device vendors . . . . .	98
4.4	Classification performance metrics (Oktoberfest experiment) . . . . .	109
		189



*List of Tables*

4.5	Crowd density classification performance metrics (Kaiserslautern public viewing event) . . . . .	129
5.1	Data set statistics (Zurich participatory experiment) . . . . .	150
5.2	Crowd density estimation approximation error (Zurich participatory experiment) . . . . .	158
5.3	Evaluating impact of the number of participatory scanners per area (Zurich participatory experiment) . . . . .	159

# Bibliography

- [1] Naeim Abedi, Ashish Bhaskar, Edward Chung, and Marc Miska. Assessment of antenna characteristic effects on pedestrian and cyclists travel-time estimation based on bluetooth and wifi mac addresses. *Transportation Research Part C Emerging Technologies*, 60:124–141, 2015.
- [2] Adnan Ghazi Abuarafah, Mohamed Osama Khozium, and Essam AbdRabou. Real-time crowd monitoring using infrared thermal video sequences. *Journal of American Science*, 8(3):133–140, 2012.
- [3] Antonio Albiol, Maria Julia Silla, Alberto Albiol, and Jose Manuel Mossi. Video analysis using corner motion statistics. In *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*, pages 31–38, 2009.
- [4] IJ Amin, Andrew J Taylor, F Junejo, A Al-Habaibeh, and Robert M Parkin. Automated people-counting by using low-resolution infrared and visual cameras. *Measurement*, 41(6):589–599, 2008.
- [5] Maria Andersson, Joakim Rydell, and Jorgen Ahlberg. Estimation of crowd behavior using sensor networks and sensor fusion. In *Information Fusion, 2009. FUSION'09. 12th International Conference on*, pages 396–403. IEEE, 2009.
- [6] Ernesto L Andrade, Scott Blunsden, and Robert B Fisher. Hidden markov models for optical flow analysis in crowds. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 1, pages 460–463. IEEE, 2006.

## Bibliography

- [7] I. B. A. Ang, F. Dilys Salim, and M. Hamilton. Human occupancy recognition with multivariate ambient sensors. In *2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, pages 1–6, March 2016.
- [8] Apple Inc. About the security content of ios 8. <https://support.apple.com/en-us/HT201395>, June 2014.
- [9] Constantine A Balanis. *Antenna theory: analysis and design*. John Wiley & Sons, 2016.
- [10] Marco V Barbera, Alessandro Epasto, Alessandro Mei, Vasile C Perta, and Julinda Stefa. Signals from the crowd: uncovering social relationships through smartphone probes. In *Proceedings of the 2013 conference on Internet measurement conference*, pages 265–276. ACM, 2013.
- [11] U. Blanke, G. Troster, T. Franke, and P. Lukowicz. Capturing crowd dynamics at large scale events using participatory gps-localization. In *Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2014 IEEE Ninth International Conference on*, pages 1–7, April 2014.
- [12] BLIP-Systems. Queue and flow management. <http://www.blipsystems.com>, 2017.
- [13] Bluetooth SIG. Bluetooth specification. [https://www.bluetooth.org/docman/handlers/DownloadDoc.ashx?doc\\_id=40560](https://www.bluetooth.org/docman/handlers/DownloadDoc.ashx?doc_id=40560), 2017.
- [14] B. Bonne, A. Barzan, P. Quax, and W. Lamotte. Wifipi: Involuntary tracking of visitors at mass events. In *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2013 IEEE 14th International Symposium and Workshops on a*, pages 1–6, June 2013.
- [15] Jeffrey A Burke, D Estrin, Mark Hansen, Andrew Parker, Nithya Ramanathan, Sasank Reddy, and Mani B Srivastava. Participatory sensing, 2006.
- [16] A.T. Campbell, S.B. Eisenman, N.D. Lane, E. Miluzzo, R.A. Peterson, Hong Lu, Xiao Zheng, M. Musolesi, K. Fodor, and Gahng-Seop Ahn. The rise of

- people-centric sensing. *Internet Computing, IEEE*, 12(4):12–21, july-aug. 2008.
- [17] Aaron Carroll, Gernot Heiser, et al. An analysis of power consumption in a smartphone. In *USENIX annual technical conference*, volume 14, pages 21–21. Boston, MA, 2010.
- [18] Antoni B Chan, Zhang-Sheng John Liang, and Nuno Vasconcelos. Privacy preserving crowd monitoring: Counting people without people models or tracking. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–7. IEEE, 2008.
- [19] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):15, 2009.
- [20] Thou-Ho Chen, Tsong-Yi Chen, and Zhi-Xian Chen. An intelligent people-flow counting method for passing through a gate. In *Robotics, Automation and Mechatronics, 2006 IEEE Conference on*, pages 1–6. IEEE, 2006.
- [21] Tsong-Yi Chen, Chao-Ho Chen, Da-Jinn Wang, and Yi-Li Kuo. A people counting system based on face-detection. In *Genetic and Evolutionary Computing (ICGEC), 2010 Fourth International Conference on*, pages 699–702. IEEE, 2010.
- [22] Jingyuan Cheng, Bo Zhou, Mathhias Sundholm, and Paul Lukowicz. Smart chair: What can simple pressure sensors under the chairs’ legs tell us about user activity. In *UBICOMM13: The Seventh International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*, 2013.
- [23] J. W. Choi, S. H. Cho, Y. S. Kim, N. J. Kim, S. S. Kwon, and J. S. Shim. A counting sensor for inbound and outbound people using ir-uwb radar sensors. In *2016 IEEE Sensors Applications Symposium (SAS)*, pages 1–5, April 2016.
- [24] A. Coşkun, A. Kara, M. Parlaktuna, M. Ozkan, and O. Parlaktuna. People counting system by using kinect sensor. In *2015 International Symposium on Innovations in Intelligent SysTems and Applications (INISTA)*, pages 1–7, Sept 2015.

## Bibliography

- [25] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 886–893. IEEE, 2005.
- [26] T. Damarla, M. Oispuu, M. Schikora, and W. Koch. Tracking and counting multiple people using distributed seismic sensors. In *2016 19th International Conference on Information Fusion (FUSION)*, pages 1593–1599, July 2016.
- [27] Anthony C Davies, Jia Hong Yin, and Sergio A Velastin. Crowd monitoring using image processing. *Electronics & Communication Engineering Journal*, 7(1):37–47, 1995.
- [28] Han Ding, Jinsong Han, Alex X. Liu, Jizhong Zhao, Panlong Yang, Wei Xi, and Zhiping Jiang. Human object estimation via backscattered radio frequency signal. In *Computer Communications (INFOCOM), 2015 IEEE Conference on*, pages 1652–1660, April 2015.
- [29] Prabal Dutta, Paul M Aoki, Neil Kumar, Alan Mainwaring, Chris Myers, Wesley Willett, and Allison Woodruff. Common sense: participatory urban sensing using a network of handheld air quality monitors. In *Proceedings of the 7th ACM conference on embedded networked sensor systems*, pages 349–350. ACM, 2009.
- [30] Nathan Eagle and Alex (Sandy) Pentland. Reality mining: sensing complex social systems. *Personal Ubiquitous Comput.*, 10:255–268, March 2006.
- [31] Nathan Eagle, Alex Sandy Pentland, and David Lazer. Inferring friendship network structure by using mobile phone data. *Proceedings of the national academy of sciences*, 106(36):15274–15278, 2009.
- [32] Evocount. Digitale besucherzählung. <https://evocount.de>, 2017.
- [33] Z Fang, SM Lo, and JA Lu. On the relationship between crowd density and movement velocity. *Fire Safety Journal*, 38(3):271–283, 2003.
- [34] Katayoun Farrahi and Daniel Gatica-Perez. Discovering routines from large-scale human locations using probabilistic topic models. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(1):3, 2011.

- [35] Julien Freudiger. How talkative is your mobile device?: An experimental study of wi-fi probe requests. In *Proceedings of the 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks*, WiSec '15, pages 8:1–8:6, New York, NY, USA, 2015. ACM.
- [36] Yuki Fukuzaki, Masahiro Mochizuki, Kazuya Murao, and Nobuhiko Nishio. A pedestrian flow analysis system using wi-fi packet sensors to a real environment. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, UbiComp '14 Adjunct, pages 721–730, New York, NY, USA, 2014. ACM.
- [37] Daniel Halperin, Wenjun Hu, Anmol Sheth, and David Wetherall. Tool release: gathering 802.11 n traces with channel state information. *ACM SIGCOMM Computer Communication Review*, 41(1):53–53, 2011.
- [38] Marcus Handte, Muhammad Umer Iqbal, Stephan Wagner, Wolfgang Apolinarski, Pedro José Marrón, Eva Maria Muñoz Navarro, Santiago Martinez, Sara Izquierdo Barthelemy, and Mario González Fernández. Crowd density estimation for public transport vehicles. In *EDBT/ICDT Workshops*, pages 315–322, 2014.
- [39] Tristan Henderson, David Kotz, and Ilya Ayzov. The changing usage of a mature campus-wide wireless network. *Computer Networks*, 52(14):2690–2712, 2008.
- [40] Marion Hermersdorf, Heli Nyholm, Jukka Salminen, Henry Tirri, Jukka Perkiö, and Ville Tuulos. Sensing in rich bluetooth environments. In *In Proceedings of WSW'06 at SenSys'06*. *SenSys'06*, 2006.
- [41] Kei Hiroi, Yoichi Shinoda, and Nobuo Kawaguchi. A better positioning with ble tag by rssi compensation through crowd density estimation. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, UbiComp '16, pages 831–840, New York, NY, USA, 2016. ACM.
- [42] Haroon Idrees, Imran Saleemi, Cody Seibert, and Mubarak Shah. Multi-source multi-scale counting in extremely dense crowd images. In *Computer*

*Bibliography*

- Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 2547–2554. IEEE, 2013.
- [43] IEEE. Ieee standard for information technology–telecommunications and information exchange between systems local and metropolitan area networks–specific requirements part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications. *IEEE Std 802.11-2012 (Revision of IEEE Std 802.11-2007)*, pages 1–2793, March 2012.
- [44] IEEE. Ieee std. 802.11n-2009: Enhancements for higher throughput. <http://www.ieee802.org>, 2017.
- [45] Irisys. Infrared people counting. <http://www.irisys.net>, 2017.
- [46] Herbert Jacobs. To count a crowd. *Columbia Journalism Review*, 6(1):37, 1967.
- [47] Ming Jiang, Jingcheng Huang, Xingqi Wang, Jingfan Tang, and Chunming Wu. An approach for crowd density and crowd size estimation. *Journal of Softwar*, 9(3):757–762, 3 2014.
- [48] Michael J Jones and Daniel Snow. Pedestrian detection using boosted features over many frames. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4. IEEE, 2008.
- [49] Pravein Govindan Kannan, Seshadri Padmanabha Venkatagiri, Mun Choon Chan, Akhihebbal L. Ananda, and Li-Shiuan Peh. Low cost crowd counting using audio tones. In *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems, SenSys '12*, pages 155–168, New York, NY, USA, 2012. ACM.
- [50] Saad Khan and Mubarak Shah. A multiview approach to tracking people in crowded scenes using a planar homography constraint. *Computer Vision–ECCV 2006*, pages 133–146, 2006.
- [51] W. Z. Khan, Y. Xiang, M. Y. Aalsalem, and Q. Arshad. Mobile phone sensing systems: A survey. *IEEE Communications Surveys Tutorials*, 15(1):402–427, First 2013.

- [52] V. Kostakos. Using bluetooth to capture passenger trips on public transport buses. *arXiv*, 806, 2008.
- [53] V. Kostakos, T. Camacho, and C. Mantero. Wireless detection of end-to-end passenger trips on public transport buses. In *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*, pages 1795–1800, sept. 2010.
- [54] Louis Kratz and Ko Nishino. Tracking pedestrians using local spatio-temporal motion patterns in extremely crowded scenes. *IEEE transactions on pattern analysis and machine intelligence*, 34(5):987–1002, 2012.
- [55] Barbara Krausz and Christian Bauckhage. Loveparade 2010: Automatic video analysis of a crowd disaster. *Computer Vision and Image Understanding*, 116(3):307–319, 2012.
- [56] Nicholas D Lane, Emiliano Miluzzo, Hong Lu, Daniel Peebles, Tanzeem Choudhury, and Andrew T Campbell. A survey of mobile phone sensing. *IEEE Communications magazine*, 48(9), 2010.
- [57] Jakob Eg Larsen, Piotr Sapiezynski, Arkadiusz Stopczynski, Morten Mørup, and Rasmus Theodorsen. Crowds, bluetooth, and rock’n’roll: Understanding music festival participant behavior. In *Proceedings of the 1st ACM International Workshop on Personal Data Meets Distributed Multimedia*, PDM ’13, pages 11–18, New York, NY, USA, 2013. ACM.
- [58] Kai Li, Chau Yuen, Salil S Kanhere, Kun Hu, Wei Zhang, Fan Jiang, and Xiang Liu. Senseflow: An experimental study for tracking people. *arXiv preprint arXiv:1606.03713*, 2016.
- [59] Ronghua Liang, Yuge Zhu, and Haixia Wang. Counting crowd flow based on feature points. *Neurocomputing*, 133:377–384, 2014.
- [60] Sheng-Fuu Lin, Jaw-Yeh Chen, and Hung-Xin Chao. Estimation of number of people in crowded scenes using perspective transformation. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 31(6):645–654, November 2001.



## Bibliography

- [61] James Little and Brendan O'Brien. A technical review of cisco's wi-fi-based location analytics. *Cisco White Paper*, 2013.
- [62] Lichuan Liu, Sen M Kuo, and MengChu Zhou. Virtual sensing techniques and their applications. In *Networking, Sensing and Control, 2009. ICNSC'09. International Conference on*, pages 31–36. IEEE, 2009.
- [63] S. Longo and B. Cheng. Real-time privacy preserving crowd estimation based on sensor data. In *2016 IEEE International Conference on Mobile Services (MS)*, pages 95–102, June 2016.
- [64] Ruihua Ma, Liyuan Li, Weimin Huang, and Qi Tian. On pixel count based crowd density estimation for visual surveillance. In *Cybernetics and Intelligent Systems, 2004 IEEE Conference on*, volume 1, pages 170–173. IEEE, 2004.
- [65] Vijay Mahadevan, Weixin Li, Viral Bhalodia, and Nuno Vasconcelos. Anomaly detection in crowded scenes. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 1975–1981. IEEE, 2010.
- [66] Aparecido Nilceu Marana, Marcos Antonio Cavenaghi, Roberta Spolon Ulson, and FL Drumond. Real-time crowd density estimation using images. In *International Symposium on Visual Computing*, pages 355–362. Springer, 2005.
- [67] Remy Melina. How is crowd size estimated? <http://www.livescience.com/8578-crowd-size-estimated.html>, 2010.
- [68] Prashanth Mohan, Venkata N Padmanabhan, and Ramachandran Ramjee. Nericell: rich monitoring of road and traffic conditions using mobile smart-phones. In *Proceedings of the 6th ACM conference on Embedded network sensor systems*, pages 323–336. ACM, 2008.
- [69] Taketoshi Mori, Yoshiko Suemasu, Hiroshi Noguchi, and Tomomasa Sato. Multiple people tracking by integrating distributed floor pressure sensors and rfid system. In *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, volume 6, pages 5271–5278. IEEE, 2004.

- [70] A. Morrison, M. Bell, and M. Chalmers. Visualisation of spectator activity at stadium events. In *2009 13th International Conference Information Visualisation*, pages 219–226. IEEE, 2009.
- [71] A. B. M. Musa and Jakob Eriksson. Tracking unmodified smartphones using wi-fi monitors. In *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems, SenSys '12*, pages 281–294, New York, NY, USA, 2012. ACM.
- [72] M Nakatsuka, H Iwatani, and J Katto. A study on passive crowd density estimation using wireless sensors. In *The 4th Intl. Conf. on Mobile Computing and Ubiquitous Networking (ICMU 2008)*. Citeseer, 2008.
- [73] Anirudh Natarajan, Mehul Motani, and Vikram Srinivasan. Understanding urban interactions from bluetooth phone contact traces. In *International Conference on Passive and Active Network Measurement*, pages 115–124. Springer, 2007.
- [74] Carol E Nicholson and B Roebuck. The investigation of the hillsborough disaster by the health and safety executive. *Safety Science*, 18(4):249–259, 1995.
- [75] T. Nicolai and H. Kenn. About the relationship between people and discoverable bluetooth devices in urban environments. In *Proceedings of the 4th international conference on mobile technology, applications, and systems and the 1st international symposium on Computer human interaction in mobile technology*, pages 72–78. ACM, 2007.
- [76] E. O’Neill, V. Kostakos, T. Kindberg, A. Schiek, A. Penn, D. Fraser, and T. Jones. Instrumenting the city: Developing methods for observing and understanding the digital cityscape. *UbiComp 2006: Ubiquitous Computing*, pages 315–332, 2006.
- [77] P.H. Patil and A.A. Kokil. Wifipi-tracking at mass events. In *Pervasive Computing (ICPC), 2015 International Conference on*, pages 1–4, Jan 2015.

## Bibliography

- [78] C. Phillips, D. Sicker, and D. Grunwald. A survey of wireless path loss prediction and coverage mapping methods. *IEEE Communications Surveys Tutorials*, 15(1):255–270, First 2013.
- [79] Jayaraman Ramachandran. Systems, methods, and computer program products for estimating crowd sizes using information collected from mobile devices in a wireless communications network, May 14 2013.
- [80] Mike Raybould, Trevor Mules, Elizabeth Fredline, and Renata Tomljenovic. Counting the herd. using aerial photography to estimate attendance at open events. *Event Management*, 6(1):25–32, 2000.
- [81] Alessandro Enrico Cesare Redondi, Davide Sanvito, and Matteo Cesana. Passive classification of wi-fi enabled devices. In *Proceedings of the 19th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, pages 51–58. ACM, 2016.
- [82] Jens Rittscher, Peter H Tu, and Nils Krahnstoeber. Simultaneous estimation of segmentation and shape. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 486–493. IEEE, 2005.
- [83] Lorenz Schauer, Martin Werner, and Philipp Marcus. 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 '14, pages 171–177, ICST, Brussels, Belgium, Belgium, 2014. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).
- [84] Natasha Schüll. The touch-point collective: Crowd contouring on the casino floor. *Limn*, 2, 2012.
- [85] Priyanka Sinha, Avik Ghose, and Chirabrata Bhaumik. City soundscape. In *Proceedings of the 13th Annual International Conference on Digital Government Research*, pages 298–299. ACM, 2012.

- [86] Beril Sirmacek and Peter Reinartz. Automatic crowd analysis from airborne images. In *Recent Advances in Space Technologies (RAST), 2011 5th International Conference on*, pages 116–120. IEEE, 2011.
- [87] M. Soyuturk, M.C. Bodur, A.B. Bakkal, and S. Ozturk. Estimating the number of people in a particular area using wifi. In *Signal Processing and Communications Applications Conference (SIU), 2015 23th*, pages 2541–2544, May 2015.
- [88] Hang Su, Hua Yang, and Shibao Zheng. The large-scale crowd density estimation based on effective region feature extraction method. In *Proceedings of the 10th Asian conference on Computer vision - Volume Part III, ACCV'10*, pages 302–313, Berlin, Heidelberg, 2011. Springer-Verlag.
- [89] Jing Su, Alvin Chin, Anna Popivanova, Ashvin Goel, and Eyal De Lara. User mobility for opportunistic ad-hoc networking. In *Mobile Computing Systems and Applications, 2004. WMCSA 2004. Sixth IEEE Workshop on*, pages 41–50. IEEE, 2004.
- [90] Ying-li Tian, Lisa Brown, Arun Hampapur, Max Lu, Andrew Senior, and Chiao-fe Shu. Ibm smart surveillance system (s3): event based video surveillance system with an open and extensible framework. *Machine Vision and Applications*, 19(5):315–327, 2008.
- [91] Mohan M Trivedi, Tarak L Gandhi, and Kohsia S Huang. Distributed interactive video arrays for event capture and enhanced situational awareness. *IEEE Intelligent Systems*, 20(5):58–66, 2005.
- [92] Mathias Versichele, Tijs Neutens, Matthias Delafontaine, and Nico Van de Weghe. The use of bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the ghent festivities. *Applied Geography*, 32(2):208 – 220, 2012.
- [93] F. Viani, E. Giarola, F. Robol, A. Polo, A. Lazzareschi, A. Massa, and T. Moriyama. Passive wireless localization strategies for security in large indoor areas. In *2014 IEEE Conference on Antenna Measurements Applications (CAMA)*, pages 1–3, Nov 2014.

## Bibliography

- [94] Florian Wahl, Marija Milenkovic, and Oliver Amft. A distributed pir-based approach for estimating people count in office environments. In *Computational Science and Engineering (CSE), 2012 IEEE 15th International Conference on*, pages 640–647. IEEE, 2012.
- [95] Shoko Wakamiya, Ryong Lee, and Kazutoshi Sumiya. Crowd-sourced urban life monitoring: urban area characterization based crowd behavioral patterns from twitter. In *Proceedings of the 6th International Conference on Ubiquitous Information Management and Communication*, page 26. ACM, 2012.
- [96] Weihong Wang, Jian Zhang, and Chunhua Shen. Improved human detection and classification in thermal images. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 2313–2316. IEEE, 2010.
- [97] Yan Wang, Jie Yang, Hongbo Liu, Yingying Chen, Marco Gruteser, and Richard P. Martin. Measuring human queues using wifi signals. In *Proceedings of the 19th Annual International Conference on Mobile Computing & Networking, MobiCom '13*, pages 235–238, New York, NY, USA, 2013. ACM.
- [98] James Weaver, Fakir Dawood, and Ali Alkhalidi. Modelling of radiation field of patch antenna for use in long range rfid crowd monitoring. In *Computer Applications Technology (ICCAT), 2013 International Conference on*, pages 1–4. IEEE, 2013.
- [99] Jens Weppner and Paul Lukowicz. Collaborative crowd density estimation with mobile phones. In *Second International Workshop on Sensing Applications on Mobile Phones. ACM Conference on Embedded Network Sensor Systems (SenSys-11), 9th, November 1, Seattle, USA*. Microsoft, ACM, 2011.
- [100] Jens Weppner and Paul Lukowicz. Bluetooth based collaborative crowd density estimation with mobile phones. In *Proceedings of the Eleventh Annual IEEE International Conference on Pervasive Computing and Communications (Percom 2013)*, pages 193–200. IEEE, 3 2013.
- [101] Jens Weppner, Paul Lukowicz, Ulf Blanke, and Gerhard Tröster. Participatory bluetooth scans serving as urban crowd probes. *Sensors Journal, IEEE*, 14 (12):4196–4206, Dec 2014.

- [102] Jens Weppner, Benjamin Bischke, and Paul Lukowicz. Monitoring crowd condition in public spaces by tracking mobile consumer devices with wifi interface. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct. International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp-16)*, September 12-16, Heidelberg, Germany, UbiComp '16, pages 1363–1371. ACM, 2016.
- [103] Jens Weppner, Benjamin Bischke, and Paul Lukowicz. Sensing room occupancy levels with ieee 802.11n wifi channel state information fingerprinting. *IEEE Sensors Letter*, 2017.
- [104] D.A. Westcott, D.D. Coleman, B. Miller, and P. Mackenzie. *CWAP Certified Wireless Analysis Professional Official Study Guide*. IT Pro. Wiley, 2011.
- [105] M. Wirz, T. Franke, D. Roggen, E. Mitleton-Kelly, P. Lukowicz, and G. Troster. Inferring crowd conditions from pedestrians' location traces for real-time crowd monitoring during city-scale mass gatherings. In *Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), 2012 IEEE 21st International Workshop on*, pages 367 –372, june 2012.
- [106] Martin Wirz, Tobias Franke, Daniel Roggen, Eve Mitleton-Kelly, Paul Lukowicz, and Gerhard Tröster. Probing crowd density through smartphones in city-scale mass gatherings. *EPJ Data Science*, 2(1), 2013.
- [107] Kristen Woyach, Daniele Puccinelli, and Martin Haenggi. Sensorless sensing in wireless networks: Implementation and measurements. In *Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks, 2006 4th International Symposium on*, pages 1–8. IEEE, 2006.
- [108] Chenren Xu, B. Firner, R.S. Moore, Yanyong Zhang, W. Trappe, R. Howard, Feixiong Zhang, and Ning An. Scpl: Indoor device-free multi-subject counting and localization using radio signal strength. In *Information Processing in Sensor Networks (IPSN), 2013 ACM/IEEE International Conference on*, pages 79–90, April 2013.

*Bibliography*

- [109] Yaoxuan Yuan, Chen Qiu, Wei Xi, and Jizhong Zhao. Crowd density estimation using wireless sensor networks. In *Mobile Ad-hoc and Sensor Networks (MSN), 2011 Seventh International Conference on*, pages 138–145, Dec 2011.
- [110] Zebra Technologies Corporation. White paper: Analysis of ios 8 mac randomization on locationing. <http://mpact.zebra.com/documents/iOS8-White-Paper.pdf>, June 2015.
- [111] Beibei Zhan, Dorothy N Monekosso, Paolo Remagnino, Sergio A Velastin, and Li-Qun Xu. Crowd analysis: a survey. *Machine Vision and Applications*, 19(5-6):345–357, 2008.
- [112] X. Zhang, Z. Yang, W. Sun, Y. Liu, S. Tang, K. Xing, and X. Mao. Incentives for mobile crowd sensing: A survey. *IEEE Communications Surveys Tutorials*, 18(1):54–67, Firstquarter 2016.

# Curriculum Vitae

Jens Weppner received his diploma in computer science from the University of Passau in 2010. Subsequently he was working as a research assistant in the Embedded Systems Lab at the University of Passau. In 2011 he spent a three month research internship at the Palo Alto Research Center (Xerox PARC) in California, USA. In 2012 he joined the German Research Center for Artificial Intelligence (DFKI) in Kaiserslautern, Germany. He was leading research projects including regional, federal, large European union and large industrial projects with settings on multiple continents. Among other top tier attendees he was selected by Google for the Google Glass workshop in 2013 in Zurich, Switzerland. His research work in the area of ubiquitous computing includes context recognition, crowd condition estimation, audio classification and activity recognition based on wearable platforms such as smartphone and Google Glass. He authored conference articles, journal articles, and holds a patent together with Xerox PARC.