Crowd Textures

Martella, C.

2017

Crowd Textures

From Sensing Proximity to Understanding Crowd Behavior

Claudio Martella

Ph.D. Thesis
Vrije Universiteit Amsterdam, 2017

This work was funded by the Dutch national program COMMIT/ under the project P09 EWiDS.

This work was carried out in the ASCI graduate school.
ASCI dissertation series number 364.

Usage of the painting "Exodus" on the cover is under kind permission of artist Lesley Anne Cornish.

Manuscript printed by GVO.

ISBN 978-94-6332-140-2

VRIJE UNIVERSITEIT

Crowd Textures

From Sensing Proximity to Understanding Crowd Behavior

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Exacte Wetenschappen
op vrijdag 24 februarie 2017 om 11.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Claudio Martella

geboren te Bolzano, Italië

promotor:	prof.dr.ir. M.R. van Steen
copromotor:	prof.dr.ir. H.E. Bal

Examiners:

prof.dr.ir. A. van Halteren	VU University Amsterdam, Philips Research
prof.dr.ir. P. Havinga	University of Twente
prof.dr. S. Klous	University of Amsterdam, KPMG
prof.dr. G.P. Picco	University of Trento
dr. S. Voulgaris	VU University Amsterdam

To my mother.
These days, nothing is in my head more than your words *"You can do it."*
Well, those and *"Stop picking your nose."*

Acknowledgements

A Ph.D. thesis is written to seem like the result of an unstoppable series of a-ha moments and eurekas. It could not be further from the truth. Pursuing a Ph.D. has felt more like "eating glass and staring into the abyss of death", to quote Elon Musk talking about being an entrepreneur. Yet, regardless of the sweat, blood and tears, it has been a challenging and greatly rewarding journey I could not have completed without the help and support of many.

My first acknowledgments are certainly for my supervisor Maarten van Steen, who took me on this journey when I had lost nearly all faith about obtaining a Ph.D., and most of the hope about having a healthy and constructive relationship with a boss. Maarten has proven it not only possible, but he has gone further by being a mentor and a model throughout these five years, teaching me with great passion how to conduct research and write papers, and most importantly making it a lot of fun. The support I received during the hard hits in my career and personal life was much appreciated and will not be forgotten. I would also like to thank Aart van Halteren for showing me that the industry can be a place for science and research, often giving me novel and original points of view and angles (for example, impressively predicting already five years ago I would end up working with beacons after my Ph.D.). Although we have not had the opportunity to do research together, I would like to thank Henri Bal for having my back (in particular when it was injured) during my last years at the VU.

The next necessary acknowledgments go to my family, who also on this journey has been by my side from start to end, by encouraging my decision and supporting its execution. During these years we have been through things nobody should have to, but we have shown our bond is strong.

I want to continue by thanking my paranymphs Marco Cattani and Kaveh Razavi for their job and for their friendship during these five years. You share positive attitude and a constant smile, together with an ability to think out-of-the-box of which I am very jealous. A mention is due for my paranymph *honoris causa* Ben Grass, who at times has coped with some of my worst edges without ever breaking my nose (regardless of his capability to do so – though everybody knows the liver punch is his favorite shot). It is clear to me only now that you have introduced me to boxing only to have the opportunity to punch me repeatedly without any legal consequence.

A special notice of acknowledgment, or perhaps of blame, goes to Elia Bruni and Matteo Caprari, who saw and nurtured the junior researcher that was still in me during my years in the industry after obtaining my M.Sc., eventually being partially responsible for the choice to go back to academia for a Ph.D.

The work presented in this dissertation was conducted under the umbrella of the EWiDS project. The project and the work have brought me to team up with a number of people from whom I have learnt great and many things, like Claudine Conrado, Keon Langendoen, Jie Li, Andreas Loukas, Ioannis Protonotarios, and Marco Zuniga, to mention a few. I feel lucky that my first years of work at the VU overlapped with the last of Matthew Dobson, on whose work much of mine is based. We have had long days and weekends of hacking, and they were all fun. A particular mention is due for the long-lasting collaboration (technically, beginning three days *before* the official start of my contract – that should have been a clear sign of what was about to happen to my life) with Hayley Hung and Gwenn Englebienne, and later also with Ekin Gedik and Laura Cabrera-Quiros. The nearly masochistic way in which we pushed each other from a crazy experiment to the next is beyond my understanding.

I would also like to acknowledge the many people I have spent time with at the VU, starting with the lovely colleagues I shared the office with, including Rena Bakhshi, Albana Gaba, and Suhil Yousaf, who managed to make the long hours feel a bit shorter, continuing with the people with whom I shared many meals, coffees and borrels, including Sara Magliacane, Dirk Vogt (in this case, many cigarettes as well...), Lionel Sambuc, David van Moolenbroek, and Ismail Elhelw, and concluding with the secretaresses extraordinaire Caroline Waij and Mojca Lovrencak, who were always there on a short notice to fix all the mess with my bureaucracy and so much more.

One of the reasons why it has been so gratifying to pursue a Ph.D. at the VU have been the opportunities of collaboration. I am glad for the hours spent working with Ana Lucia Varbanescu, whose passion for teaching and research have opened my eyes about the fun and responsibilities that come with supervising students, and Alexandru Iosup's team at TU Delft. The Lighthouse team meetings with Peter Boncz and Spyros Voulgaris have been something to look forward to. The collaboration for the CoBrA experiment with Jeana Frost has pushed me in the realms of the social sciences with positive surprises. I owe countless tokens of gratitude to Armando Miraglia for the number of hours (and many extra hours) put in our projects. Your dedication and work ethic is a continuous source of inspiration.

I would also like to acknowledge the many students I have had the pleasure to work with who put effort, passion and great work in our projects, including Renske Augustijn, Christian Chilipirea, Jesse Donkervliet, Sinziana Filip, Unmesh Joshi, Ian Morozan, Per Ohme, Andreea Petre, Peter Rutgers, Andreea Sandu, Dimo Stoyanov, and Vlad Tudose.

I would like to extend my gratitude to Salvatore Scellato who enabled my internship with the geo/beacons team at Google in London, UK, where I had the fantastic opportunity to apply some of my research in a legendary environment, and to the engineers I had the pleasure to work with there, including Vilius Naudziunas, Stefano Maggiolo, Nandana Dutt, Jonathan Morace, and the rest of the team.

I also had the pleasure to intern in the Big Data team at Telefonica Research in Barcelona, Spain, working together with the all-greek team including Dionysios Logothetis, Georgos Siganos and Ilias Leontiadis, a collaboration that has been fruitful beyond the time boundaries of the internship.

I am glad to acknowledge that, against common belief, the concepts I learnt during my doctoral years are useful beyond the domain of distributed computer systems. For example, for a BBQ at the beach, one should buy disposable grills in excess. They can, and will fail. Redundancy is key. I have also learnt that in most cases a Ph.D. thesis manuscript is only used to raise the height of some computer screens or to level a table. For this reason, the cover of this manuscript is printed on growth paper, meaning that it contains seeds of a mix of wild flowers. Furthermore, the manuscript is printed on biodegradable 100%-recycled paper. This way, the worst-case scenario where the manuscript is trashed in a backyard, or in nature in general, turns into a much more desirable and colorful best-case scenario. Unfortunately, the "forget-me-not" flower seeds were not available.

Finally, I would like to conclude with some apologies to Alan Turing and John von Neumann. In the acknowledgments of my B.Sc. and M.Sc. theses I gratefully noted that it was the work of the two scientists that pushed me to decide to pursue a career in Computer Science. I also suggested, this time perhaps with a tiny note of blame, that my at-the-time status of single may be re-conducted to the very same decision. Well, maybe ironically, it was at the first conference in my doctoral years that I met Vera, to whom I am grateful for her love and support, and for pushing me everyday out of my comfort zone. Jokes on me, I guess.

Claudio Martella
Rio de Janeiro, Brazil, December 2016

Publications

Design for Crowd Well-being: Needs and Design Decisions, J. Li, H. De Ridder, A. Vemeeren, C. Conrado, and C. Martella. In *Proceedings of the International Conference on Planning and Design (ICPD)*, 2013.

Design for Crowd Well-being: Current Designs, Strategies and Future Design Suggestions, J. Li, H. De Ridder, A. Vemeeren, C. Conrado, and C. Martella. In *Proceedings of the International Association of Societies of Design Research (IASDR)*, 2013.

Crowd Textures as Proximity Graphs, C. Martella, A. van Halteren, M. van Steen, C. Conrado, and J. Li In *IEEE Communications Magazine*, 2014.

How Well do Graph-Processing Platforms Perform? An Empirical Performance Evaluation and Analysis Y. Guo, M. Biczak, A. Varbanescu, A. Iosup, C. Martella, and T. Willke. In *Proceedings of the IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, 2014.

Benchmarking Graph-Processing Platforms: A Vision, Y. Guo, M. Biczak, A. Varbanescu, A. Iosup, C. Martella, and T. Willke. In *Proceedings of the ACM/SPEC International Conference on Performance Engineering, Work-in-Progress (ICPE)*, 2014.

From Proximity Sensing to Spatio-Temporal Social Graphs, C. Martella, M. Dobson, A. van Halteren, and M. van Steen. In *Proceedings of the IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2014.

Adaptive Partitioning of Large-scale Dynamic Graphs, L. Vaquero, F. Cuadrado, D. Logothetis, and C. Martella. In *Proceedings of the International Conference on Distributed Computing Systems (ICDCS)*, 2014.

Spinner: Scalable Graph Partitioning in the Cloud, C. Martella, D. Logothetis, A. Loukas, and G. Siganos. In *arXiv*, 2014.

How Was It? Exploiting Smartphone Sensing to Measure Implicit Audience Responses to Live Performances, C. Martella, E. Gedik, L. Cabrera-Quiros, G. Englebienne, and H. Hung. In *Proceedings of the ACM Multimedia Conference (MM)*, 2015.

Leveraging Proximity Sensing to Mine the Behavior of Museum Visitors, C. Martella, A. Miraglia, M. Cattani, and M. van Steen. In *Proceedings of the IEEE International Conference on Pervasive Computing and Communications (PerCom)*, 2016.

A Landscape of Crowd-Management Support: An Integrative Approach, N. Wijermans, C. Conrado, M. van Steen, C. Martella, and J. Li. In *Safety Science*, 2016.

Powerful and Efficient Bulk Shortest-Path Queries: Cypher Language Extension & Giraph Implementation, P. Rutgers, C. Martella, S. Voulgaris, and P. Boncz. In *Proceedings of the International Workshop on Graph Data Management Experiences and Systems (GRADES)*, 2016.

On Current Crowd Management Practices and the Need for Increased Situation Awareness, Prediction, and Intervention, In C. Martella, J. Li, C. Conrado, and A. Vermeeren. In *Safety Science*, 2016.

Visualizing, Clustering, and Predicting the Behavior of Museum Visitors, C. Martella, A. Miraglia, J. Frost, M. Cattani, and M. van Steen. In *Pervasive and Mobile Computing*, 2016.

Exploiting Density to Track Human Behavior in Crowded Environments, C. Martella, M. Cattani, and M. van Steen. In *IEEE Communications Magazine*, 2017.

An Open-Space Museum as a Testbed for Popularity Monitoring in Real-World Settings, M. Cattani, I.Protonotarios, C. Martella, J. van Velzen, M. Zuniga, and K. Langendoen. In *Proceedings of the International Conference on Embedded Wireless Systems and Networks (EWSN)*, 2017.

Contents

1 Introduction

> *Deciding what to do is as important as deciding what not to do.*
>
> — **Steve Jobs**

Zetabytes of user data are being collected every year, and the volume is expected to grow more than linearly in the coming years. In the data economy, data about our behavior on the Internet and the World Wide Web, including activities on social networking and social media sites, is used to compute, for example, search results, product recommendations, and news filters, not to mention to train the machine-learning models that back smart features, and to improve user interfaces through A/B testing.

The diffusion of smart phones has allowed us to interact with such platforms also when we are away from home, and their “apps” now support many of our activities, resulting in even more data being produced and collected. Today, smart phones are more powerful than the supercomputers of 50 years ago, and they are packed with all sorts of sensors and interaction interfaces. As such, they are in the privileged position to measure, together with our online activities, many aspects of our offline behavior.

Smart phones and smart watches, and the upcoming wave of wearables, are opening a window in the digital world over our lives in the physical one. And it appears to be just the beginning.

The number of ubiquitous and pervasive technologies is expected to increase with the implementation of paradigms like the Internet of Things, where all kinds of devices and appliances distributed around us will be able to collect and share signals about our behavior. The processing of such signals with new machine-learning algorithms should finally allow intelligent machines to predict, recommend, and personalize, to support our lives while receding in their background.

To manage all the data and applications we can leverage the infrastructure built during the last decades, comprising reliable operating systems, high-performance computing platforms equipped with application-specific integrated circuits, large-scale distributed stores and queues, reliable and efficient cluster technology, as well as miniaturized, powerful, and energy-efficient embedded devices.

Now, a larger attention needs to be put into finding the right signals and the sensors to collect them, together with the right models and algorithms to quantify and qualify our behaviors at massive scales. Most importantly, as many of our behaviors are social and the result of our interactions with the environment, in the future a particular emphasis is required on cohesive approaches that do not focus on the single individual or signal, but that take into account the collective, embodied, and situated nature of human behavior.

Spatio-temporal proximity is one signal, and it is the subject of this dissertation.

1.1 Characterizing spatio-temporal proximity

As context shapes the meaning of a word in a sentence, objects and people around us define who we are and what we do. Due to the physical constraints of our body and the ways we communicate, in the real world we stand physically close to the objects and the people with whom we interact. We face the people we talk to, we stand close to the appliances and objects we are using or need to reach, we share spaces with co-workers and family.

Modeling and measuring spatial proximity, and how it develops over time, is important for the understanding of human behavior.

Spatio-temporal proximity has been the object of extensive research in recent years, from the identification of co-location in the workplace to the characterization of face-to-face social interactions, to name two examples. In its simplest form, proximity can be modeled as a relationship between two entities, representing whether at a specific moment in time the two entities were within a certain physical distance from each other.

This information can be further enriched, for example, by including the exact distance between the entities, the angle (e.g., if they are facing each other), as well as a timestamp to encode the time when such relationship took place. While the information contained in a single proximity relationship is simple and minimal, these dimensions, as well as the possibly large number of relationships, allow for a number of interesting analyses.

Depending on the behavior to be measured and on the application, different variables of proximity information can be explored across different dimensions, and they can be “zoomed” in and out to focus on micro and macro aspects of the behaviors.

1.1.1 The spatial dimension

Thinking about proximity, the first dimension that comes to mind, and perhaps the most intuitive, is the spatial dimension. Different ranges of distance can capture different aspects of social behavior. For example, interactions between individuals within 5-7 meters of distance relate to the study of proxemics and the various kinds of spaces (e.g., personal space, social space etc.), whereas longer ranges of distance up to around 15 meters can be used to identify co-location in the study of cooperation (e.g., individuals sharing an office or part of an open space). In certain occasions, one can even relate to, though perhaps not sense, longer range geographical proximity (e.g., kilometers of distance), as in the case of the study of cooperation within and between organizations on a territory.

Often, when focusing on a particular type of interaction or relationship, it is not necessary to measure the actual distance between the entities and a definition of a boundary or threshold can be enough. For example, it can be sufficient to capture proximity relationships that take place within a distance of 2-3 meters (without measuring the precise distance) to study f-formations[1] and mingling behavior.

Furthermore, depending on the type of behavior, the angle of proximity can determine the type of relationship or interaction. The most clear examples are conversations and f-formations, which are characterized by a spatial and orientational relationship. To measure mingling behavior and f-formations it is common to measure so-called face-to-face proximity, that is when individuals are facing each other within an angle of some 30-60 degrees at short distance. In contrast, for other behaviors, as in the case of crowd dynamics like clogging and pedestrian lanes, it is important to capture proximity across a more uniform space around the individuals.

As with distance, it is not always necessary to characterize the actual angle of proximity between two individuals or between an individual and an object, but it is sufficient to capture whether the entities were within a certain maximum angle, as in the case of face-to-face proximity.

Different variables in the spatial dimension set different constraints to the choice of modality and technology used to sense proximity.

[1] An F-formation arises whenever two or more people sustain a spatial and orientational relationship in which the space between them is one to which they have equal, direct, and exclusive access.

1.1.2 The temporal dimension

A proximity relationship can have a temporal property that defines when the relationship was valid. Different units of time can be used to determine the duration of validity of a specific proximity relationship like, for example, a second. As in the case of the variables in the spatial dimension, also the unit of time is usually chosen depending on the underlying behavior, and in particular the expected dynamicity or rate of change. For example, to identify co-location a time unit of some minutes may be enough, while for highly dynamic scenarios like crowd dynamics one may need more fine grained time units, like seconds.

Clearly, the choice of time unit is influenced by the constraints dictated by the technology used to sense proximity, as higher sampling frequencies tend to be more expensive to manage, for example, due to higher battery consumption.

Furthermore, temporal properties can be used to aggregate a number of proximity relationships over an interval of time (e.g., to filter and aggregate proximity relationships for a given day), in scenarios where the quantity of time spent in proximity is relevant to the study of the behavior. Other examples of temporal aggregations are techniques based on sliding windows used, for example, to compute how proximity changes over time. Finally, the temporal information can be used to investigate consequentiality, as in the case of the study of epidemics, where the temporal dimension is used to follow how information (or diseases, mood, behavior etc.) spreads based on spatial proximity.

1.2 Sensing modalities

There is a number of different approaches and techniques to capture proximity, which can be divided in two groups.

1.2.1 Tracking absolute location

On one hand, *absolute* location of individuals and objects is tracked at all times (e.g., as 3D or latitude-longitude coordinates), and proximity relationships are captured by computing distances and angles (assuming one can also track who a person is facing) between these absolute coordinates, for example, through the Euclidean distance metric. Examples of such approaches are video cameras, the Global Positioning System (GPS), and indoor localization systems (ILS). These approaches require deploying a fixed infrastructure as well as a mobile infrastructure, as in the case of GPS and ILS where a device needs to be worn by the individuals.

While techniques based on cameras do not need individuals to wear a sensor, they tend to be less reliable when operating in challenging settings with obstructions, changes in lighting conditions and requiring merging multiple points of view. Furthermore, they are known to operate poorly with high density of individuals, and need the deployment of a large number of cameras to track large crowds. Finally, computer-vision algorithms tend to be compute-intensive and may cause concerns regarding the need to collect privacy-invasive footage.

Conversely, GPS and ILS techniques tend to be cheaper to compute and easier to anonymize (e.g., devices can be associated with unknown identities for many applications), but still require the deployment of a landmark infrastructure that may be unfeasible to install at a scale at required granularity. Furthermore, GPS and ILS are also known to operate poorly in high-density conditions, and usually provide localization errors of some meters in the best case. Finally, the capturing of angle of proximity often depends on additional sensing modalities, as, for example, through the use of a compass.

1.2.2 Tracking relative location

On the other hand, *relative* location of individuals and objects is tracked as binary proximity relationships, with the only information that is recorded being whether two entities, including individuals, objects, and points-of-interest, were close to one another (hence, relative) at a given time. Like GPS and ILS systems, this approach requires individuals and objects to be instrumented, but this time the sensors must detect other sensors only within a certain distance and angle.

The advantage of this approach is a limited dependency, if any, on fixed infrastructure and landmarks, as no intermediate step of absolute localization is necessary. Furthermore, this approach is particularly advantageous as sensors can be installed at and worn only by entities of interest, and the detection of proximity between two entities depends solely on the two sensors involved. This way, it is simpler to deploy a large-scale network of proximity sensors, with fewer centralized components, and sensors can be added only where and when needed, making this approach suitable to real-world dynamic deployments both outdoors and indoors (e.g., in festivals, city events, but also in museums, train stations etc.).

Typical technologies used to implement proximity sensors are radio-based (e.g., Bluetooth low-energy (BLE) and Zigbee), ultrasound, and infra-red sensors, where a unique identifier can be transmitted and received within a certain distance range. Angles of detection can be enforced either by means of directional antennas, or by leveraging the shielding effect of the body of the individuals when the sensor is worn, for example, on the chest. Recently the industry has provided some attempts of standardizing a protocol for proximity sensing through small and inexpensive BLE transmitters and BLE-enabled receivers (e.g., smart phones and smart watches). Examples of such efforts are Apple's *iBeacon* and Google's *Eddystone*.

1.3 Proposed approach

The work I present in this dissertation follows the second approach of capturing and modeling social behavior by leveraging relative-proximity information, and in particular by means of radio-based proximity sensors.

The main research question around which this work develops regards whether social behavior can be measured with proximity sensors, and whether by analyzing proximity data one can gain a better understanding of the measured behavior, and produce insights and feedback to ensure the safety and comfort of a crowd.

1.3.1 Crowds, crowd behavior, and crowd textures

Social behavior is a loose term that is intentionally used in this dissertation to refer to a wide range of behaviors involving a number of individuals. Proximity information can be used to study a large number of different social behaviors in a variety of contexts and applications that range from the patterns emerging in a crowded place, the interactions in a workplace, the dynamics developing in a large city, to the foundations of other technologies like the Internet of Things, to name a few examples. However, it is outside the scope of this dissertation to characterize and define all of them.

Instead, I focus on one instance, that is crowds and crowd behavior.

The term *crowd* is used to refer to a large number of individuals, and it can be characterized differently depending on the object and field of study. In this dissertation, I use the term to refer to a number of individuals gathering in the same place over a well-defined period of time, which can last from hours to days, with all members not necessarily being in that same place at the same time, nor necessarily sharing the same goal or interest.

Such definition fits crowds gathering at train stations, music festivals, museum exhibitions, city parades etc., and it differs from definitions of crowds where a number of individuals are dislocated and distributed geographically in very different places and may never be in the same place at the same time (e.g., when one wants to study the profile and behavior of customers of some coffee chain or the means and patterns of commute in a city, like in mobile crowd sensing studies).

Crowd behavior can be characterized according to a number of different aspects, all related to proximity information.

First, there are purely spatio-temporal aspects, as in the case of so-called crowd dynamics, which refer to bi-directional pedestrian lanes, flows, queues, clogging and congestions, etc. Second, there are more social aspects of crowd behavior, where related and interacting individuals spend time close to each other, for example, while mingling or otherwise. Third, there are other aspects regarding relationships between individuals with respect to profile, taste, and behavior, which can be characterized by how similar people tend to distribute their time in a similar way not only with other people, but also at certain places and interacting with certain objects. Finally, there are higher-level aspects of crowd behavior that include, for example, emotions, moods, and actions, which are more suited to be studied by means of other types of sensors (e.g., microphones, accelerometers, video cameras, galvanic skin response sensor, etc.), but where proximity information can still play a role, for example, to study how these elements spread in the crowd.

In this dissertation, I argue that it is not possible to understand the behavior of a crowd by focusing on the behavior of the single individuals in isolation.

In fact, similar to a flock of birds, the behavior of the individuals in a crowd can be understood only if their nearby individuals are taken into account. Note that objects and points-of-interest can act as proxies to extend the definition of *nearby*, by allowing to relate individuals that were at the same place at different moments (e.g., in front of a painting, in a bar, in a specific room). This notion of dependency on nearby individuals creates a recursive relationship, where the members of a crowd are interleaved by a series of proximity relationships.

I coined the term *crowd texture* to refer to the series of interleaved proximity relationships characterizing a crowd and its behavior.

I also argue that different crowd behaviors are characterized by different crowd textures (or that different crowd behaviors affect the crowd texture in a different way), and that by mining the series of measured proximity relationships crowd behaviors can be recognized and identified. While, intuitively, one can think of recognizing crowd behavior in the texture of a crowd as performing pattern matching on a series of spatio-temporal proximity relationships, in practice, a broad range of data-analysis algorithms and techniques can be used.

Crowd textures, and the so-called proximity graphs used to represent them, are the elements that bind together the chapters of this dissertation.

1.3.2 Problem statement

Answering the research question presented in this dissertation requires finding a solution for the following problems.

- Current crowd-management practices and tools are lacking certain information and capabilities, which existing technologies are unable to provide or fail to provide in certain conditions. We want to understand, pulling directly from the professionals operating on the field, requirements and needs of crowd managers to help them manage safer and more comfortable crowds.
- It is not clear what are the properties and limitations of relative localization, and how such type of information can be leveraged to model and identify crowd behavior. We want to identify effective models to represent the texture of a crowd, starting from proximity information, that can enable the identification of the underlying crowd behaviors and dynamics.
- Sensing relative proximity with radio-based sensors is unreliable due to the inherent limitations of wireless sensors networks and radio-based communication. We want to design an architecture of a system to collect proximity information in the real-world that can operate in the face of the challenges dictated by scale, density, and extreme environmental conditions.
- Once a reliable measurement of a crowd texture is performed, it is not clear how to analyze it to identify, quantify, or qualify the underlying crowd behaviors. We want to design a set of algorithms, without assuming a perfect and complete measurement, to understand what is happening in the crowd. Moreover, the algorithms should operate in respect of the privacy of the members of the crowd.

- Identifying and characterizing crowd behavior are the first steps towards ensuring the safety and comfort of a crowd. We want to study strategies to gain insights and turn them into actionable information. This may comprise feedback to the crowd managers and/or to the crowd.
- Finally, proximity information has its limitation, perhaps missing higher-level aspects of crowd behavior. We want to understand how proximity information can work in cooperation with other sensing modalities to study the facets of crowd behavior that relate more with the inner experience of the individuals.

1.4 Outline and contributions

This dissertation tries to tackle the above problems and makes the follow contributions.

In Chapter 2, we study current crowd management practices by means of ten interviews conducted with professionals managing some of the largest crowds in The Netherlands. Needs and requirements are identified, including the limitations of current approaches and technologies. In particular, the interviewees reported a strong need for increased situation awareness, prediction, and intervention, positioning spatio-temporal aspects of crowd behavior, like density, flows, and congestions, at the top of their priorities. Recommendations are provided within the framework of a techno-social system for crowd management comprising two feedback-control loops.

In Chapter 3, we dive into the definitions and properties of crowd textures and proximity graphs. Examples of analyses to identify pedestrian lanes, bottlenecks, and social groups are provided, along with a discussion about the importance of taking into account the context around a crowd, as well as advantages of instrumenting objects as well as individuals with proximity sensors. A real-world use case is presented, by identifying social groups within face-to-face proximity from data collected during an ICT conference.

In Chapter 4, we analyze some of the sources of unreliability in mobile and fixed proximity sensing when applied to crowd monitoring. A filter with a few parameters, and a technique to choose good values for them, is presented and evaluated both in simulation and in a number of real-world experiments. The technique can double the sensitivity of the measurement, reconstructing nearly all missed proximity detections and ignoring spurious ones.

In Chapter 5, we focus on tracking an indoor crowd inside a museum, studying in particular person-to-object face-to-face proximity relationships. A pipeline of filters is presented, comprising a particle filter adapted to directional sensing, together with a number of applications to visualize, cluster, and predict the behavior of the museum visitors, with data collected during a real-world 5-days experiment conducted at the CoBrA museum of Amsterdam. We also look into the results of a number of interviews and a focus group conducted with museum staff regarding the effectiveness of the visualizations.

In Chapter 6, we look into an extension of the previous technique that takes into account also person-to-person proximity, to overcome limitations caused by high densities. The technique is evaluated during a large-scale experiment with nearly a thousand participants conducted at the NEMO museum, a particularly challenging setup due to the peculiar features of the NEMO building. Together with the filtering pipeline, an analysis to detect crowded events as well as flows between floors is presented.

In Chapter 7, we study how proximity sensors can be used together with accelerometers to study the response of a crowd to a dance performance. An experiment is presented where audience behavior was tracked before, during, and after the performance, to try to identify the response of the audience (e.g., enjoyment) by looking into the collective movements of the (seated) audience during the performance, and the differences in mingling behavior exhibited before and after the performance.

The dissertation terminates with a brief discussion and conclusions.

2 On current crowd management practices and the need for increased situation awareness, prediction, and intervention

If the user can't use it, it doesn't work.

— **Susan Dray**

2.1 Introduction

In the previous chapter we have discussed how spatio-temporal information can help the study of social behavior. We have also specified the particular focus of this dissertation on crowd behavior as one instance of social behavior. To begin the work, in this chapter we first focus on gaining a better understanding of the state-of-the-art practices and technologies used by the professionals that work with crowds and their behavior on a daily basis, that is crowd managers.

Crowd management is essentially a set of collaborative practices between a number of different actors, e.g., event planners and managers, emergency services, local authorities, transport authorities, stewards, and the crowd itself [29, 148]. These practices start months ahead of an event. In fact, as we discuss in this chapter, preparations take about 90% of the efforts. Usually a multi-agency approach is followed, incorporating all relevant parties, to enable a wide range of knowledge and expertise to be drawn upon. Preparation activities include detailed risk analyses to identify and prioritize potential risks, use and development of comprehensive "what-if" scenarios to consider management strategies and contingency plans, establishment of a control point to

The contents of this chapter have been originally published in **"On current crowd management practices and the need for increased situation awareness, prediction, and intervention"** C. Martella, J. Li, C. Conrado, A. Vermeeren - *Safety Science* 2016, and have been slightly modified to improve readability.

coordinate all activities and personnel. The remaining 10% consists of implementing the plan, comprising monitoring crowd activity to identify potential problems, and intervention, that in extreme conditions can result in crowd control. It must be noted that the focus of crowd management is facilitating crowd activities, hence proactively preventing, or quickly resolving, problems. The correct and effective execution of such practices is crucial to the success of an event, with the most important outcome being the safety and comfort of the crowd [2, 51].

It has been argued that a more systematic approach to crowd management could have avoided recent accidents in large crowded events [41, 28]. We postulate that new developments in technology, including mobile sensors, decision-support systems, and novel communication and interaction paradigms, can support crowd management operations during the planning and implementation of an event. However, as also supported by our results, currently the success of operations is still mainly dependent on the personal experience and skills of the crowd management team, with little or no aid from technology.

Towards a better understanding of the limitations and requirements of current crowd management practices, in particular regarding the role of technology, we present the perspective of crowd managers. We interviewed 10 crowd managers daily involved with managing large crowds, including a stadium hosting tens of thousands of visitors, a large train station, a multi-day music festival, a yearly celebration involving more than a million people. A main result emerging from our interviews is that crowd managers feel the need for instruments offering an increased situation awareness, a more reliable and timely monitoring of the state of the crowd, and the ability to predict and steer the behavior of the crowd without use of force.

In this chapter, we make the following contributions. First, we present background and supporting literature, including crowd behavior modeling and prediction, mobile sensing, and decision-support systems. Second, we present current crowd management practices, as they emerged during our interviews. In particular, we focus on the role of technology and its limitations. Third, we present crowd managers' requirements for future technology to support their operations. Finally, we discuss opportunities and recommendations within the framework of a techno-social system.

2.2 Background

A generally accepted definition of a crowd is that it is a large gathering of diverse people at the same physical location, at the same time, not necessarily sharing the same goal or interest [147]. Understanding the behavior of crowds, and how to manage them effectively, is still a scattered effort that involves different fields including theoretical physics, sociology, psychology, computational science, and artificial intelligence. Recently, studies have been published with overviews of common crowd management practices [29, 65], but more work is required. The literature about crowds and crowd behavior focuses on theoretical modeling of the psychology of crowd behavior [126, 117], predicting crowd behavior through physics-inspired models, recognizing behaviour through various kinds of sensors and analysis.

An approach to studying crowd behavior is by synthesizing it through crowd behavior prediction models. Crowd behavior prediction models are also used for a-priori planning of events through simulation [132, 165, 7, 125]. A popular example of a crowd behavior model is the social-force model [67]. The models usually target so-called crowd dynamics, referring to patterns of crowd movement, and more precisely to “the coordinated movement of a large number of individuals to which a semantically relevant meaning can be attributed, depending on the respective application” [120]. Examples include a queue of people, the formation of uni-directional “lanes” in bi-directional pedestrian flows, the intersection of these lanes, or a group of people at a specific location. Approaches to crowd modeling and simulation have been extensively surveyed [139, 15, 48].

A different approach is to investigate how to detect and recognize crowd behavior. Computer-vision techniques have been employed to characterize and automatically detect anomalies in a crowd [166, 161]. The diffusion of pervasive and ubiquitous technologies such as smart phones and smart watches, has enabled the monitoring of social behavior through a wide range of sensing modalities, from temperature, to movement, to spatial proximity [141, 11]. For example, smart phones have been used to detect crowd dynamics such as pedestrian flows and bottlenecks, and social groups [153, 154, 156]. In particular, crowd dynamics such as pedestrian lanes and clogging have a strong spatio-temporal nature that can be captured as so-called crowd textures using proximity sensors [97] (see Chapter 3). Accelerometers can be used to characterize queues, and activities such as running and walking [83]. Finally, microphones can be used to measure the mood of a crowd [33] or recognize locations and places [84][1]. Some of these approaches are grouped also under the term Ambient Intelligence (AmI), referring to "electronic systems that are sensitive and responsive to the presence of people" including context and social-aware miniaturized pervasive computing devices and sensors, which can be envisioned to enhance and support, for example, crowd monitoring and evacuation [104].

While synthesizing and recognizing crowd behavior has been addressed in the literature, less attention has been dedicated to how such data can help crowd managers make effective decisions in the control room, for example, during an event. Existing works either tend to focus on managing disasters and emergencies [23, 111, 92, 114, 10, 73], or on very specific cases such as air traffic control [17, 95] and underground stations [134, 93], overlooking how technology can be used to support decisions *before* accidents happen during an event, or to support planning and debriefs.

Theories on socio-technical systems recommend new systems to be designed and operated with a holistic approach that optimizes both technical and social factors [31, 30, 35, 34]. This body of work is crucial to the design of systems that make use of technology to support the work of crowd managers. While these principles have been applied to the domain of technology and work design over the last decades, a broader and braver approach is necessary to extend their reach, for example, to crowd management [40]. In this chapter, we take a technological stand within this attempt, by studying how technology currently helps (or fails to help) crowd managers in their practices, and how existing and new research can serve the work of crowd managers in organizing and managing safer and more secure crowds.

2.3 Methodology

In this section we present our participants and the methods used to conduct the interviews and analyze the collected data.

[1]Note that these techniques differ from the emerging field of Mobile Crowd Sensing (MCS) [60]. MCS uses mobile devices to collect information from individuals dislocated and distributed in wide areas, and defines a crowd as a large number of individuals that may be distributed geographically in *different locations* (and even different countries), or that visit the same location at *different times*.

Crowd expert	Description	Crowd size	Crowd duration
P1 Indoor Music Festival	Chief organizer of an annual indoor music festival, coordinating the preparation, registration, staff training, communication during the festival	2000	6 hours
P2 Indoor Conference	Chief organizer of an annual large indoor conference, coordinating the registration, communication, transportation, parking, catering etc.	1000	12 hours
P3 Central Train Station	Crowd manager of a central train station in a capital city, managing the crowds in daily situations and in large events	250.000	4 hours
P4 Police	Crowd manager involved with large crowds e.g., on a national festival	700.000	8 hours
P5 Security Company	Head of a security company, consulting on organization and management of various crowd events	1000-100.000	hours to days
P6 Barrier Company	Head of a barrier company designing and building custom barriers for various types of crowd events as well as managing their layout	1000-100.000	hours to days
P7 Outdoor Music Festival	Manager of an annual outdoor music festival, coordinating the site construction, ticketing, crowd flow control, transportation, parking, catering etc.	60.000-100.000	3 days
P8 Stadium	Crowd manager of a stadium, managing the crowds for various events, such as concerts and football matches	55.000	4-5 hours
P9 Theme Park	Manager of a theme park,focusing on managing the daily crowd flows, queues and large crowds (e.g., crowds watching a music fountain) during holidays	40.000-60.000	12 hours
P10 Train Station Flow	Crowd manager of a central train station, monitoring real-time crowd flows via video cameras, Bluetooth and Wi-Fi signals	180.000	12 hours

Tab. 2.1 Description of the experts interviewed.

2.3.1 Participants

We carried out 10 individual interviews with 10 crowd experts. We selected and approached 10 organizations in The Netherlands known for hosting and managing among the largest crowds in the country. From each organization, we interviewed a senior professional with experience in dealing with large crowds. Type of event, location, visitors profiles, time of year, among others, define different scenarios of crowd behavior and the different strategies to manage them. For this reason, we chose organizations that allowed us to cover the widest range of events and crowds, from those emerging at peak hours in train stations to those in multi-day outdoor music festivals. Note that also within the same type of location, e.g., a train station, experts manage diverse scenarios. For example, train stations must deal with both week-day peak-hours crowds and day-long special celebration events, with hundreds of thousands of people coming in from all over the country. We summarize the participants and their domain of expertise in Table 2.1. We also included an expert from the Research & Development department of an organization specialized in designing and building barriers for large events, such as music festivals and parades. As such, he presented a different perspective of the requirements and the use cases of the crowd managers. Finally, the organization we dubbed "Security Company" differs from the other organizations due to their consultancy-oriented business model, that includes the delivery of crowd management trainings and workshops, as well as consulting on events organization and management. While diverse, the profile of the crowds managed by the participants matches our definition of a crowd provided above, that is of a large number of individuals gathering at the same location at the same time [147].

2.3.2 Interview process

We chose to conduct semi-structured interviews because they guarantee consistency in the topics that were addressed, but also freedom for the participants to diverge and provide their unique and personal perspective when necessary. The approach prevented interviews falling into a strict question-response pattern and encouraged the raise of theme-related new questions that can adequately elicit the issues to compose a more comprehensive report. Each interview was scheduled around 1 to 1.5 hours at the work place of the participant, and it was recorded with a voice recorder. All the artifacts accessed in the interviews, for example,

Fig. 2.1 (Left) An example of statement card with a quote. (Right) The magnetic wall with the categorized statements.

sketches, booklets, photos and maps, were collected at the end. We began the interviews by posing questions about three themes: (i) daily operations, (ii) crowds characterization, and (iii) use of technology. Then, we triggered the participants to develop further each theme by talking about concrete stories, rather than about general and abstract concepts.

2.3.3 Data analysis

We analyzed the data following a creative on-the-wall method [121] consisting of four steps: (i) transcription, (ii) interpretation, (iii) categorization, and (iv) presentation.

We started by transcribing and timestamping the interviews. The artifacts were used to aid the process and were included in the corpus. After the transcriptions, a team of three researchers coded the text of each interview as follows. First, each researcher independently selected relevant paragraphs. Then, the team collaboratively grouped overlapping choices into statements cards. If no consensus could be found, the researchers would either decide to discard the paragraphs, or to do another pass on the transcriptions. A statement card consisted of a statement and a group of selected quotes cut out directly from the printed transcripts.

At the end of the process, the three researchers generated 241 statement cards in total. For the following session, a fourth researcher joined the team. To categorize the statements cards, the four researchers followed a process resembling a bottom-up clustering process. Statement cards were grouped inductively into categories, and so were the resulting categories themselves, when possible, until no more categories could

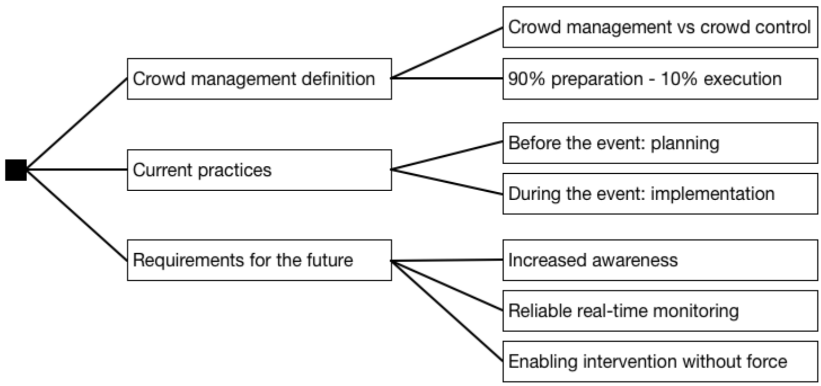


Fig. 2.2 The hierarchy of categories generated by the bottom-up clustering of the statements. Each category further contains common themes emerged during the interviews.

be generated or grouped. Figure 2.2 presents the first two levels of the hierarchy of categories, which we use to present our results in Section 2.4. The clustering was not directed by any pre-defined hypothesis, and each category name emerged during the process. The session was carried out in a room with walls covered by magnetic white boards. A dozen of A1 white paper sheets were fixed on the wall with magnets. The statement cards with relevant information were put together on the same A1 sheet. Figure 2.1 shows an example of a statement card and a part of the magnetic wall.

The findings were presented in the form of a poster[2]. As we discuss in the next session, time is an important dimension in the management of crowds. For this reason, the visualization is constructed around a time-line. The poster visualizes the categories together with the most important statements. Interesting quotes are printed with larger font sizes to guide the attention towards the more detailed summaries. We present a detailed analysis of the results in the next section.

2.4 Findings

In this section we present our findings, organized following the hierarchy of categories pictured in Figure 2.2, emerged from the data analysis presented in Section 2.3.

[2]The poster can be downloaded from `https://goo.gl/O38B9F`

2.4.1 Overview: on the definition of crowd management

All experts strongly emphasized two main distinctions during the interviews. The first distinction related to the difference between *crowd management* and other practices like *crowd control* and *riot control*. The second distinction related to the two phases that constitute the crowd management practices, namely what happens *before* the event and what happens *during* the event. We organize our findings around these distinctions.

Crowd management is usually defined as the set of measures taken in the normal process of facilitating the movement and enjoyment of people [18], for instance measures to control the distribution of people over a certain area. This definition fits that of the interviewees. From their responses, crowd management is taken to refer mostly to the preparations for a given event and it involves predicting what is going to happen and preparing for it, i.e., designing for the desired behavior of the crowd. The preparations involve all aspects, namely getting people into the site, people participating to the event, and getting people out of the site. These preparations usually start much ahead of the event, e.g., six months or more. The resulting plan or design for a given event concerns the technical infrastructure and operational measures needed for the safety, well-being and enjoyment of the crowd. The effort to produce the plan was estimated by the interviewees to be about 90% of the total effort for the event's crowd management. The remaining 10% refers to the (mostly) operational measures, including potentially emergencies during the event itself, which implement and support the plan for the event.

Crowd management was distinguished from crowd control. The latter includes all measures taken once crowds are beginning to or have gone out of control. In other words, crowd management is proactive while crowd control is reactive [18]. This distinction is reflected both by the uneven allocation of resources towards preparation, and by an emphasis of monitoring, predicting, and steering behavior during the event to avoid the need for crowd control. In this chapter we focus in particular on current practices, limitations and requirements of crowd management.

2.4.2 Current practices

We start by presenting current practices. Where not specified, crowd managers did not mention use of technology, or explicitly reported none.

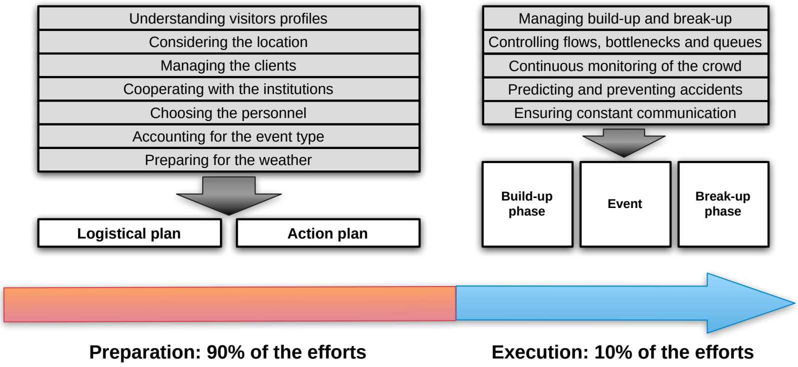


Fig. 2.3 Crowd management can be divided in two parts: the planning of the event, that takes up to around 90% of the efforts, and the execution of the plan during the event, taking about the remaining 10% of the efforts. Multiple elements must be taken into account when generating the scenarios and the plans. During the event most efforts concentrate on managing the different phases of the event, and predicting and avoiding certain behaviors, more than securing the crowd once these happen.

Figure 2.3 summarizes the elements that comprise the current practices, including both planning and execution of the event.

Before the event: planning

The type of knowledge required to produce a plan in crowd management includes (i) expert knowledge based on experience, (ii) guidelines learnt from past events' data, (iii) trial and error, (iv) common-sense knowledge, as well as (v) computer simulations of the crowds in the event. Planning is typically done within a management team, in which each member has his or her responsibilities and disciplines. They input their knowledge, expertise and experience on past events to the management process. Typical roles and responsibilities include (i) the transportation to the event, which can start far away from the event location, (ii) the security, sometimes taken care by or in collaboration with the police, (iii) the barriers built on the location to contain or steer the crowd, (iv) the event manager, taking care of the event plan and representing the various stakeholders.

There is no general recipe to produce the plan, and a number of factors need to be taken into account, including (i) visitor profiles, (ii) location, (iii) client, (iv) institutions, (v) personnel, (vi) event type and (vii) weather. We now proceed with a description of the content of the crowd management plan, and later in this section we describe the aforementioned factors in detail.

The planning starts with a definition of the desired behavior the crowd management team wants to obtain from the crowd. The content of the plan is the outcome of all the decisions that should eventually steer the crowd towards that desired behavior during the event. Such plan is generally composed of two parts.

The first part is a *logistical plan* that concerns decisions about, for example, the number of tickets sold, the mobility plan and the resulting layout of the event site including the position of barriers, entrances, exits, toilettes, bars, the transportation systems, the provisioning of food and beverages. The main goal of these decisions is to allow the crowd to move freely and safely, but at the same time avoid certain dangerous or unpleasant situations caused, for example, by uneven distribution of the crowd, obstructions, bottlenecks, and dissatisfaction. Communication with the crowd is also taken into account. Hence, the plan includes decisions, for example, about signs, screens, event program, and maps. Additionally, the plan contains information about the number of individuals in the personnel and their profiles, including their protocols and briefings.

The second part resembles an *action plan* which consists of a number of "what-if" scenarios and strategies regarding how to respond to each given situation. This includes the preparation of scenarios for several alternative "normal" situations, depending on weather, type of public, locations, most likely crowded areas, peak hours, as well as for dealing with emergencies, including evacuation plans and crowd control. Scenarios are typically constructed with the help of expert knowledge and computer simulations, and the planning also indicates the courses of action that may be taken in the given situation. Critical scenarios in events involving large crowds are the arrival and departure of the people at and from the site, so special attention must be paid for the preparations for these moments.

Understanding the visitors is the first step in event planning. All the planning for an event has to be adapted to its visitor profiles. The age and gender of the visitors are of great importance. It is easy to imagine that young, aggressive and male adults in football stadiums are more difficult to manage than less active and well-behaved older adults. The visitors' familiarity or former experience of an event also plays an influential role. For example, it is common to see loyal visitors to some annual events, and this influences their behavior.

When the visitors mainly consist of groups of friends or family members, e.g., at a theme park, strategies for keeping them together should be considered in the planning with particular importance. The transportation of the visitors should also be taken into account, making sure there are enough parking places or clear paths connecting the public transportation to the event field. This again may depend on the visitor profile, as more mature visitors may visit the event through their car, while an event attended by teenagers may require better planning of public transportation.

The location is evaluated next. Event locations can have various, sometimes very specific characteristics. An indoor event must follow stricter rules than the outdoor ones. For example, the amount of visitors allowed to an indoor event is limited by the amount of emergency exits. On the other hand, outdoor environments tend to have fewer regulations, as present fewer physical constraints that can limit the behavior of the visitors in case of emergencies, such as, for example, walls, gates, and stairs.

Besides whether the event takes place indoor or outdoor, a location-related aspect that requires particular attention is the level of mobility, namely whether the crowd is seated, standing, or continuously walking. Some events may include multiple levels of mobility. For example, in a conference, people are mostly sitting during the presentations, walking around during breaks and standing to listen to a scholar explaining his/her posters. Several managers pointed out that managing the seated crowds, e.g., in a stadium, is much easier than dealing with the randomly moving crowds, e.g., in a train station's hall or when people are approaching the event location from multiple points.

Is the location built on grass or concrete? This is the third consideration related to location. If the event is on an outdoor grass field, more attention will be paid to the weather resistant measures. For example, by preparing the sawdust for soaking up the water in case of rain. The fourth consideration is whether the event is in city center, in a neighborhood, or in a tourist attraction. Organizing the event in the limited area of the city center is less regulated than an event in a neighborhood, because a city center is designed for social activities while a residential neighborhood is less tolerant for noise.

An important part of running a safe event is *managing the clients*. The client of an event includes artists that perform and organizations that promote a festival. Sometimes, crowd management reasons can influence the choice of clients and/or the design of schedules. For example, for a multi-stage music festival, inviting a very dominant and famous band to perform may produce dangerous skewed distributions of visitors across stages. For this reason big bands are often scheduled at the same time on different stages. So, planning may also need to consider the behavior of certain clients or the behavior they may cause. Some clients may behave in unplanned or expected way, creating dangerous situations in the crowd by, for example, attracting many individuals or creating excitement in areas not designed to handle such conditions.

Part of managing a crowd is *cooperating with the institutions*. Various institutions can be promoters or owners of an event, like the government. However, often governmental institutions are not experts in crowd management. They promulgate regulations and give permits to event organizers, or fine the organizers due to noise, damages to the locations, and so on. They work mainly as a partner or stakeholders in crowd management, who can provide support or help coordinate in an event. For example, the manager of the security company, the manager of a central train station and the manager of a stadium, all believed that the police partly belongs to the government, whose responsibilities are different from those of security personnel in the event. Hence, institutions can act as resources, but also as constraints in the collaborative work of creating a crowd management plan.

Choosing the personnel by hiring the right amount of people with the right set of skills is a significant part of the preparation. Certain personnel works well for one event, but may not adapt to another. For example, personnel you need to manage the audience of a television show is different from the people needed in a football game. In a football game, crowds can be very large, and sometimes aggressive. Plus, one needs more personnel for ticket sales on the location. Part of the personnel are the security stewards and the first-aid staff, focusing on safety issues. The catering professionals take charge of the food and beverage that is considered as a big element contributing to visitors' happiness. The critical role that skills and communication play in the choice, instruction, and management of the personnel, shows once again the importance of the collaborative nature of the work.

Many *attributes of an event* have impact on planning. How do you control the crowd size in a ticket-less event? What are the different strategies in a heavy-metal concert and in an classical concert? What is the duration of the event? All these questions raise the concerns on the internal attributes of an event. The goal of an event sometimes includes making profits. Finding a balance between maximizing profits while keeping the crowd safe is one of the biggest challenges. For example, if an event is free for all the visitors, how many visitors will come may be unknown in advance. A free party is planned differently from a paid party. For this reason, when possible, free events are still organized with tickets to control the size of the crowd.

Finally, *preparing for the weather* is very important as it can change the conditions of the event very quickly. The weather mainly affects outdoor events. Some events plan also for bad weather, and need to guarantee that the temporary architectures, such as tents and decorations, resist also to bad weather conditions. Weather can also largely influence transportation. A storm may drive a large crowd of visitors of an outdoor carnival to the train station in a very short time, potentially paralyzing the train station.

During the event: implementing the plan

During the event, the role of the crowd management team is to assess the condition of the crowd, evaluate the active scenarios, predict future scenarios developments, and execute the related actions according to the plan. Given the limited range of actions a crowd management team can execute during the event without resulting in crowd control, many of the strategies implemented by crowd managers concentrate on avoiding density reaching critical levels, more than actually reacting to it. For this reason, many of these strategies focus on particular moments and areas of the event, e.g., the entrances and exits, the locations where queues can form, e.g., shops and toilettes. In this section we present important lessons stressed by the experts.

Managing the build-up and break-up phases of the event, such as the few hours preceding and following the event, often involves different strategies and considerations. As a general strategy, the management of the crowd begins as far as possible from the site, guiding the gathering in a safe and comfortable condition. Although not always possible, guiding the crowd as early as possible provides managers with a wider time window to predict future developments, and allows for proactive actions. For example, depending on the size, the type of event, the location, and the actors involved, the management of the crowd can start from the public transport stations, the neighboring towns, up to the extra-regional areas.

When possible, multiple routes and entrances should be made available to the crowd. This often depends on the location. For example, modern stadiums and urban areas can have routes that start already from dedicated motorway exits. When such infrastructure is not in place, use of barriers, turnstiles, and signs should help the formation of these routes. The type of event has a relevant impact on this phase. For example, a long lasting event without a fixed main attraction, such as a festival or a national celebration, will present a fairly more diluted and continuous flow of individuals over the event duration, compared to a football game or a concert.

Central to the management of the safe movement of the crowd is *controlling of pedestrian flows, bottlenecks, and queues*. Ensuring the emergence of safe pedestrian flows and queues that do not develop into bottlenecks and clogging behavior is rated among the highest priorities of both phases of crowd management. Many factors affect the movement and flows of pedestrians, and their characterization is central to their understanding and control to ensure the safety of the crowd [129, 127]. Crowd and pedestrian dynamics do not only play a a role in the build-up and break-up phases, when the crowd arrive at and leave the site, but also throughout the whole duration of the event. For example, flows and queues can generate also from state to stage between concerts, or between platforms in a train station.

In general, three main strategic guidelines are applicable to the scenarios of flows, queues, and bottlenecks: (1) keep the flow moving, (2) avoid long intervals of times where the individuals are forced to wait still (it is generally accepted that waiting for longer than 8 minutes may affect the mood of the individuals in a queue), (3) keep the individuals informed of waiting times, the causes of the block, and the condition of the crowd in front of them. The strategies to obtain a continuous and safe flow of pedestrians range from a good combination of capacity planning and human resources, to communication (including signs), and site design (i.e., with the aid of barriers). For example, a simple yet practical strategy is the avoidance of money exchange at the food stands, in favor of particular coins or prepaid cards to be bought in advance that minimize transaction times. At a theme park, a queue can be entertained by the surrounding attractions.

Barriers can be used to divide the crowd in smaller and more manageable groups, to guide people towards exits and entrances, to support queuing, to create routes and detours, to avoid the stagnation of individuals in certain areas, such as gates or corridors. While barriers can be positioned to “mold” the crowd in specific areas, it is sometimes necessary also to temporarily remove them, for example, in train stations when very large crowds are expected for special events. In those occasions, when the site is close to the maximum capacity, barriers can turn into dangerous obstacles.

Communication is of utmost importance, as it is used to keep the crowd informed about the decisions made by the managers, and to support independent decisions by the individuals. Communication is also used within the management team to exchange information about monitored areas, to brief agents/stewards about plan changes or actions to be taken. To a certain extent, communication is one of the few and most powerful means the managers have to influence and steer the behavior of a crowd without use of force.

As far as communication to the crowd is concerned, the content of the communication can range from the densities in the different areas, early information about public transport, time schedules, different path options, and changes to the schedule or weather conditions. Within the constraints of crowd management, the role of communication is to discourage the crowd to move towards certain areas of the site, and persuade them to take different routes, sometimes also representing longer detours that allow the mass to spread more evenly. Communication can be supported by physical infrastructure such as screens, barriers, and signs. It is recommended to spread them evenly across the site to reach the largest audience, and position them in places that can host potentially large groups to avoid bottlenecks. For example, screens with train schedules can sometimes be installed already outside of the train station. When the event allows it, information about paths and routes can be spread already in neighboring urban centers through fliers, radio and television broadcasts, Internet sites and social media. These means of communication can complement more traditional ones, such as loud speaking and megaphones.

On the other hand, communication within the team has different goals. First, it allows to share information about the state of the crowd, such as the distribution of people in different areas, the formation of flows, warnings about anomalies, or more logistical information such as the need for specific resources. This type of information generally travels from the agents in the field to the control room, where it is processed and used for decision making. Moreover, communication is needed to provide agents and stewards in the field with actions to execute as a result of the decisions made in the control room. Finally, communication is used to coordinate actions on the field, among agents and stewards. Communication within the team occurs through different channels and technologies, depending on the endpoints, the density, and the distance to cover. For example, for short ad hoc communications between two people in a sparse area, telephones can be sufficient. However, cellular mobile phones have problems functioning in high-density situations. Also, the C2000[3] can be used when actors working for different institutions are involved, and in particular for the management of emergencies. Walkie-talkie and other radio-based communication tools can be used to broadcast information from one point to multiples at the same time. These instruments often support multiple channels, so that communication endpoints can be multiplexed and information overload can be avoided.

Monitoring of the state of a crowd is currently human-centric. Information about the state of the crowd is collected by stewards and agents in the crowd, through heuristics and visual estimations. When technologies like video cameras are used, they are monitored by humans in the control room. In other words, they allow for centralization of information, but the information is still processed by humans. UAVs (Unmanned Aerial Vehicles) can be equipped with cameras: they can fly over areas, used to monitor in detail queues, spot riots, and detect abnormal situations. These remotely controlled systems provide valuable information, but they are often not legally permitted due to the risk imposed on the crowd, in case of malfunctioning. Automatic processing of video streams is still not wide-spread among crowd management practitioners, and still present low accuracy at high densities.

In principle, automatic monitoring of a crowd is possible via barriers. Turnstiles can count the amount of people that entered the event site, and their speed can be contained from the control room to control flows. Also, barriers can incorporate pressure sensors to monitor the state of the crowd in critical points e.g., in front of a stage. Nonetheless, this information is usually used to validate and design the barriers layout for future events.

Recently, social media, such as Twitter and Facebook, have been used to monitor the use of certain keywords to detect emergencies and feedback from the crowd. Moreover, mobile phones provide a good source of information as they can be used to approximate counts of individuals. This can be usually achieved either by counting the number of telephones registered to a cell, or, for example, by counting the number of telephones with active Bluetooth connectivity. However, these types of radio-based systems operate poorly in highly dense scenarios.

[3]The C2000 is a private radio-based encrypted communication infrastructure used by dutch emergency services for public safety.

An important aspect of crowd monitoring is timely information exchange and integration between the different actors. Often, different agencies such as the police, the municipality, the national railways, collect information about the crowd that can be useful for the other actors. For example, data collected from the police about flows directed towards the train station can be of great value for those managing the crowd at the station. Exchange of information is currently performed on a face-to-face basis in the control room or through the phone. Finally, also external factors should be monitored as part of the process, as they have an impact on the crowd and the event. Weather conditions, for example, can influence highly the schedule of the event, forcing people to leave the event in advance, or even representing a danger in itself in the most extreme cases.

2.4.3 Limitations of current practices and crowd managers' requirements for the future

In this section we discuss the challenges and limitations of current practices, and the requirements for the future as they were identified and outlined by the crowd managers. In Table 2.2 we summarize the requirements for the future together with the current limitations of crowd management practices.

Increasing situation awareness and decision-making support

For what concerns the planning phase, most crowd managers acknowledged that predicting all possible and relevant situations that may develop is an essential component of crowd management, yet far from trivial. In the scenario planning performed during the preparations for an event, for instance, biases towards specific scenarios (for example too optimistic or too pessimistic scenarios) often exist, even amongst experienced crowd experts. Moreover, it may be difficult to take into account situations that the experts have never experienced before. In addition, computer simulations of crowd behavior, sometimes used in these predictions, do not fully capture all essential mechanisms that are relevant in a given setting. Hence, crowd managers require more support for the generation of sound and comprehensive plans.

Regarding the implementation of the plan, crowd managers emphasized the need for increased situation awareness. That is, crowd managers require to be informed in advance about how a certain situation

Requirement	Current limitations
Increased situation awareness and decision-making support	• limited, biased, and over-simplified what-if scenarios • inability to generate unforeseen conditions • unvalidated and unrealistic computer simulations • inaccurate estimation of crowd future states • poor overview of the crowd (density, movement, flows) • limited support of decision-making support systems
Reliable real-time monitoring and communication	• human-dependent and limited monitoring (e.g., on-the-field stewards) • surveillance cameras are not ubiquitous and mostly human-operated • data is collected not in real-time • coarse-grained and unreliable data with high densities • communication happens verbally and through few shared channels • little sharing of collected data between actors
Enabling intervention without use of force	• few non-pervasive means to communicate with the crowd • inability to provide timely preventive feedback to the crowd • only fixed screens and loudspeakers are available • infrastructure such as barriers and gates are passive with little control

Tab. 2.2 Current practices' limitations and requirements for the future.

is developing, to help them act accordingly and as early as possible. For instance, they would like to have an estimate of how soon an area will get overcrowded and how early to give feedback to the crowd and scouts to handle the situation. Essential parts of situation awareness are perception, comprehension, and prediction [52]. To increase the former, most crowd managers expressed the need to have a spatial overview of the crowd, including information such as density, movement and flows in real-time. Density in particular can affect the flow and speed of movement of the crowd [57] and can hence convey relevant information to the crowd managers about the current and future state of the crowd. Such a dynamic "map" of a crowd could provide situation awareness to crowd managers and should be available at three different levels, namely (i) specific areas of interest, e.g., bottleneck areas that may get overcrowded such as entrances / exits, corridors, and stairs, (ii) the whole area where the crowd is present, for instance the whole area of a football stadium that is used by the crowd, (iii) areas around / outside the total area of interest, for instance the parking areas around / leading to the football stadium. Ideally, the map would include additional information about the crowd, such as the profiles of the people, their mood, the placement of exists, stewards and other infrastructure. To increase prediction, the information in the map needs to include trends in crowd behavior and dynamics, and how they develop over time.

Reliable real-time monitoring and communication

Crowd managers expressed the need for reliable means to measure the state of the crowd, that can operate particularly in critical conditions. Current "instruments" to survey large crowds are the stewards on the ground inside the crowd and surveillance cameras watching over the crowd. These approaches are limited as they may lead to personnel missing relevant information. Feeds from the surveillance cameras are monitored by personnel who cannot watch all screens at all times. Current automatic approaches, like computer-vision image analysis of video streams or automatic counting from mobile phones traces, do not operate at high density and large-scale. Finally, crowd managers desired to collect and integrate information currently stored in "silos" from different sources. For example, smart turnstiles and barriers are used to collect information about the crowd for future planning. Crowd managers emphasized the importance of obtaining and using this information in real-time during the event.

Similarly, crowd managers expressed need for reliable communication within the team as well as with the crowd in critical conditions. Current approaches present limitations similar to those regarding monitoring, that is they are mostly human-dependent and manually operated, causing information loss or overload. For what concerns communication to the crowd, managers stressed the need for effective means to communicate beyond current loud-speakers and screens.

Enabling interventions without usage of force

Concerning the actions that must be taken to manage the crowds, crowd managers expressed the need for mechanisms to effectively control the movements of the crowd. In particular, a mechanism was sought which would allow the coordination of the movements of groups of people in a crowd. Conceptually, a "traffic light", in the own terms of one of the managers, that would control which groups of people could move, where and when. Moreover, crowd managers expressed the need for means to evenly distribute the crowd in a given space, e.g., a train platform or a concert area, as well as to evenly distribute the crowds in time, for example, to prevent that large numbers of people arrive to and depart from an event all at the same time. Solutions for queue management in large crowds were also mentioned as important requirements, which could support “flexible queuing”, for example, allowing people to leave and re-join a queue, or even “queue-less” events and help the speeding up of queues.

While some of these interventions could be implemented through new physical and mechanical infrastructure, crowd managers acknowledged that in order to take actions towards the crowd, communication with the crowd is of utmost importance. Therefore crowd managers expressed the need to communicate well with the crowds. Mechanisms are thus needed which can very clearly and effectively provide the crowds with information such as: (i) overview of the crowd situation (ii) reasons for any troubles and delays, (iii) predictions, waiting times, (iv) advices, alternatives, orders.

2.5 Recommendations: the anatomy of a techno-social system for crowd management

In this section we provide a systematic discussion on how technology may improve the support of crowd management practices. We frame our recommendations around a *techno-social system* comprising of two feedback control loops. If we consider the interaction between the crowd management team and the crowd as a social system, then we can define a techno-social system for crowd management as a system where technology supports and augments such interaction. We envision a system implementing data-driven practices common to other industries. Examples of such practices are (i) collection, processing and sharing of large streams of heterogeneous data coming from mobile and fixed sensors, (ii) support for analysis and sense-making of crowd data, for example, by means of tools for exploration and visualization of large datasets, and (iii) recommendation and implementation of intervention strategies executed for example through actuators, such as turnstiles and barriers, and communication. At the same time, we envision empowered individuals in the crowd taking autonomous decisions based on feedback provided by the system.

A number of different technologies and state-of-the-art techniques can fit into such framework, and could aid tackling the limitations and drawbacks presented in Table 2.2. In Table 2.3 we provide some hypothetical concrete examples of how the implementation of our model would impact on the operations outlined by the crowd managers (note how the operations in Table 2.3 match the requirement in Table 2.2). We go in the details in the following sections.

2.5.1 Crowd management as two feedback control loops

In a nutshell, a feedback control loop comprises three phases: (i) *measurement* of input, (ii) *control* of input, and (iii) actuation through *feedback*. Crowd management can be seen as two of such loops. One that happens during the event, and one that takes place between events. Looking at crowd management practices as feedback control loops allows us to reason about how technology can help, in all three phases, the group work of crowd managers towards a more effective and data-driven approach. Figure 2.4 summarizes the relationship between the two loops and their operations.

The loop within each event

The first feedback control loop takes place during the event, and comprises all the operations to (i) collect information about the state of the crowd, (ii) assess and predict the state of the crowd within the planned scenarios, (iii) intervene, if necessary, to prevent uncomfortable and dangerous situations. The loops should be operated at short intervals (e.g., seconds or few minutes), to ensure real-time monitoring and quick response.

The measurement phase is perhaps the phase that has received most recent attention in the literature. Existing pervasive and ubiquitous sensing infrastructure have been proposed and investigated as a means to measure the movement of flows of people, the areas affected by undesired crowd mood, but also the capacity of parking lots and the development of weather conditions. This infrastructure includes mobile sensors such as smart phones and RFID-enabled bracelets used as festival tickets, as well as, for example, fixed sensors like video cameras and weather stations. However, crowd managers expressed a need for real-time measurement of crowd behavior, and a dissatisfaction with the unreliability of current solutions at high densities. Recent work on

Operations	Current	Future
Crowd monitoring and communication	Data about the crowd is estimated by the agents on the field and behind the surveillance cameras screens, and it is communicated verbally to the control room. Data is managed and shared manually by the different actors operating in the control room.	Data from fixed and mobile sensors, such as video cameras, smart phones, and e-bracelets is collected, shared, and processed automatically to estimate in real-time densities, flows, congestions, crowd mood, and further supports automatic decision-making support systems and prediction models.
Planning and decision-making support	What-if scenarios, and logistical and action plans are generated by the managers based on personal skills and know-how, and sometimes simulated with synthetic models. During the event, decisions are made based on the information communicated by the agents on the field, by matching the static what-if scenarios with the current state.	Historical data collected from sensors and other events is used to support the generation of what-if scenarios and plans, and to simulate different conditions. The models are validated with past data. During the event, the data collected in real-time is used not only to match the current what-if scenario, but also to enhance the overview of the crowd managers about the current state of the crowd and the predictions about future outcomes.
Intervention without use of force	Information is communicated verbally from the control room to the agent on the field who operate with the crowd and provide feedback either verbally, aided by loudspeakers, or through screens. Gates and barriers are operated manually in response to the different orders given by the control room.	Feedback is computed automatically and transmitted to the devices of both the agents and the members of crowd, as well as to the fixed screens. The decisions of the crowd managers result in actuations by the smart turnstiles, gates and barriers, which help the agents steer the crowd.

Tab. 2.3 Overview of the impact of the envisioned techno-social system on the crowd management operations emphasized by the crowd managers as requiring aid from technology.

decentralized protocols, for example, to detect density [25], hold the promise of providing reliability at large scale.

Once a continuous stream of data coming from these sensors would be established, it can be shared among agencies, and processed automat-

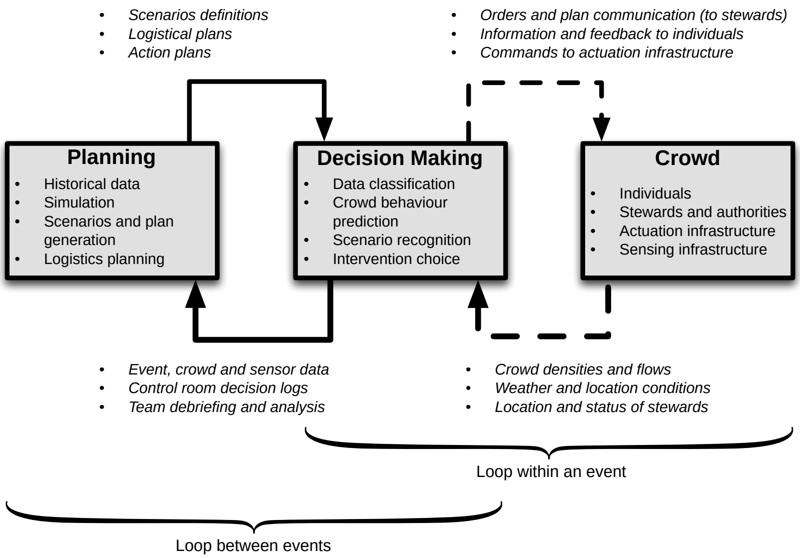


Fig. 2.4 Crowd management can be seen as two feedback loops. The first loop happens within an event and it is characterized by (i) data collection from personnel and sensors (ii) decision making in the control room and on the field based on the data, the scenarios prepared and the plans, (iii) feedback in the form of information, orders, and commands to the infrastructure. This loop is executed at high frequency to ensure quick response-time to changing conditions and emergencies. The second loop happens between events and it allows the planning of the following events to be adapted based on the lessons learnt during the previous (i.e., based on collected data during the event, debriefs, expert analysis, and decision-support systems).

ically to help sense-making about the behavior of the crowd. This is the control phase of the loop. Crowd managers have expressed a desire for instruments that can increase situation awareness and support decision making. Unfortunately, most current prediction models about crowd behavior, while numerous, still lack evaluation with real data [27]. As also reported by one of our interviewees, they are still too inaccurate to be useful. Increase in collection of crowd data can be an opportunity for improvement in validation of current models, and aid the design of novel and more comprehensive models [64, 15]. Adaptive and interactive applications can help crowd managers identifying current active "what-if" scenarios by recognizing and highlighting relevant pieces of crowd state, and recommending related interventions [36]. Four interviewees reported the need to visualize the flows of people within the location on a map, and the predicted densities for the near future. Visualization can be a powerful tool to facilitate computer-supported cooperative work [75]. Current work on visualization of mobility traces within urban areas and large scale festivals provide examples of such interfaces [101, 152].

Regarding the feedback phase, again smart phones and other wearable devices are an opportunity for crowd managers to communicate with crowd members, and platforms to support autonomous decisions by the individuals in the crowd. For example, apps running on smart phones can allow fine-grained location-aware communication targeted to specific zones of the event location with personalized information [155]. While current music festival apps allow to share and visualize the location of friends on a map, they do not display, for example, the flows of people between stages, or the densities around the different spots. Such information, automatically provided by the system, could allow individuals to make informed decision, and avoid unpleasant situations [19]. With respect to fixed infrastructure, barriers are also opportunities to communicate information about the state of the crowd, for example, through a displays showing densities at the front of a queue or near an exit, or by indicating alternative routes through lights.

The loop between events

The second feedback control loop consists of the process of crowd management itself, and includes (i) debriefing outcomes of a previous event, (ii) analyzing possible flaws in the plan or in the decision making process, and (iii) updating of the planning processes to accommodate new lessons.

The utility of the information collected during an event does not finish at the end of the event. The databases of sensed data, together with the log of decisions made by the management team during the event, provide a solid base for a systematic reasoning about the outcome of the event [89]. Interviewees underlined the generation of the what-if scenarios as being at risk of bias, lack of coverage and depth. A more data-driven approach can reduce errors and bias, as well as allow experts to recognize new scenario that happened during th event. Scenarios can be recommended to the crowd managers based on the data collected over time, during previous events. Moreover, such scenarios can be automatically characterized by estimates based on historical data. Also, data can facilitate the reviewing procedures at the end of the event [65]. Finally, sensed data can be fed into crowd behavior simulators, to overcome the aforementioned limitations related to absence of rich scenarios and validation [64, 59], or fed into computational models to support operational research techniques for task force deployment [47]. The result is a live process, where data-driven approaches support collaboration within the crowd management team from the phase of planning to the phase of implementation, during and between events.

2.5.2 Implications

In this section we discuss implications and constraints that need to be considered when introducing more technology within such a techno-social system.

Keeping the human in the loop

While we advocate a bigger role of technology in crowd management, we emphasize that such transition shall happen with the human, both as a member of the crowd management team and of the crowd, in the center. When designing crowd applications, one should take into account explicit and implicit motivational factors. For instance, users in the crowd may require incentives to share their data and, for example, utilize an energy-consuming application on their phone. Providing a platform to help locate friends could be an example of an incentive to share with the system current locations and social ties. Moreover, individuals have different psychological needs when it comes to well-being [91]. Crowd members seek for fulfilling higher level of psychological needs, i.e., staying autonomous, connected, competent and respected in normal crowd conditions, while their focus will immediately change to low level needs, i.e., safety and security issues, when unexpected things happen.

For what concerns crowd managers, it is important to support their planning and collaborative decision-making, in particular in emergency situations [77]. Interviewees pointed out that current technological solutions often lack comprehensiveness or accuracy, which resulted in rejection of the system. Perceived usefulness of an instrument is known to be strongly influencing the intentions of the user, even more than perceived ease [39]. Importantly, the goal of decision-support system should be to aid the expert decision of a crowd manager, and not substitute it [122].

Guaranteeing privacy to the crowd

While not directly mentioned by the interviewees, ensuring privacy and trust are critical adoption barriers of such a system. For example, detection of flows and bottlenecks as well as mood, require constant probing for sensitive information such as location and emotions. Moreover, privacy is user specific, that is each individual has a different perception of privacy. Various techniques have been proposed to solve the issue of privacy in sensing, and they include anonymization [135], cryptographic techniques [160, 54], and data obfuscation [5, 61]. Crowd data, in particular with respect to crowd dynamics, lends itself to these techniques. For example, the characterization of pedestrian flows and their volumes, does not require knowledge about the identity of the comprising users.

Minimizing the need for new infrastructure

Introducing technology can often imply new and expensive investments. Two distinctions are necessary in this case. First, there are gatherings that take place in fixed locations, designed for the purpose of hosting crowds. Examples are train stations, music halls, stadiums, and theaters. These locations already employ substantial infrastructure, starting from barriers, video cameras, turnstiles, sometimes Bluetooth and Wifi scanners. In these scenarios, the interviewees reported being open to the installation of new infrastructure, given a reliable functioning in critical conditions.

Second, there are events that take place in locations that are not designed to host large crowds. For example, periodic parades and celebrations taking place in city centers and neighborhoods. In these occasions, it is difficult and expensive to temporarily deploy new infrastructures pervasively across town. As reported by the interviewees, smart phones are widely adopted and minimize the need for additional infrastructure. Moreover, event planners often cooperate with telecom companies to obtain cellular data about densities. Finally, additional cellular towers can sometimes be installed to alleviate the problems of congestion.

2.6 Discussion

We interviewed 10 crowd managers to gain an understanding of the current role of technology in their current practices, including limitations and requirements. We have formulated our recommendations within a techno-social system framework. Despite the fact that the 10 volunteers agreed on many of the fundamental definitions about their practices and needs, regardless of their diverse domains of work, it is likely that our results are influenced by different biases, for example, cultural, organizational, and geographical, as all the managers operate in The Netherlands. To be able to generalize the results of the interviews more work needs to be done in different geographical, cultural and economical contexts. Nonetheless, the consistency of our results with the other work [29, 65, 18] suggests that our results can be generalizable at least to similar contexts to those in the Netherlands.

While the volunteers are indeed experts with decades of experience, we should not treat the data completely as “ground truth”. To validate some of the statements and assumptions, further on-ground and observational work should complement the insights extracted from the interviews. We suggest this work to be conducted both in the control room and on the field, during the different phases of the planning and implementation of an event.

It is unclear whether many of the technological problems and limitations we described in this chapter are specific to crowd management, and how much of the related research should be adapted to in the context of crowd management specifically. For example, radio-based communication is known to operate poorly at high densities and hence it does not represent a crowd-specific problem. Yet, crowds are not characterized solely by high densities but also by movement patterns that are innate and hence exploited, for example, by so-called epidemic protocols used by wireless sensors to disseminate data and communicate at large. At the same time, the bodies of individuals influence communication in a much different way than metallic objects as, for example, cars, making some of the related work on transportation systems not directly applicable to crowds.

The vast majority of existing literature on crowd and pedestrian behavior focuses on modeling and predicting behavior. More recent work has focused on how to devise sensing infrastructure to collect real-time data about the behavior of a crowd, and support sense-making. Less attention has been dedicated to how to support crowd management specific decision making and even less to how to intervene and steer crowd behavior to ensure the safety of the individuals. Tackling these challenges requires multi-disciplinary work that comprises investigating how to influence the behavior of a crowd, how to choose the right action to obtain a certain reaction from the crowd, what information to communicate to the crowd and with which medium and feedback. Answers to these questions are far from trivial and, to this date, mostly unknown.

2.7 Conclusions

We have presented crowd management as a set of collaborative practices. The successful management of an event depends on the cooperation and communication between these actors, and the crowd. Our findings show that crowd managers do already look at technology to solve some of their problems, and see it as an opportunity for future developments. As technology plays a bigger role in crowd management, a number of technical and technological challenges need to be tackled, from different fields and practices. We have addressed these challenges within the framework of a techno-social system. This work shows that there is space for more support from technology at different stages of the planning and implementation of an event, and it motivates and suggests new directions for research to support the safe management of crowds.

3 Crowd textures as proximity graphs

> *Perfection is reached not when there is nothing left to add, but when there is nothing left to take away.*
>
> — **Antoine de Saint Exupéry**

3.1 Introduction

In the previous chapter we have seen that crowd managers are lacking proper support from technology when it comes to situation awareness, prediction, and intervention. The core behaviors pointed out by the interviewees are behaviors with a strong spatio-temporal nature, in particular so-called crowd dynamics. In this dissertation, I argue that relative-proximity information can be leveraged to help researchers and practitioners understanding, among other social behaviors, crowd dynamics, towards filling the gaps identified by the crowd managers.

The term "crowd dynamics" is used to refer to patterns of crowd movement, and more precisely to "the coordinated movement of a large number of individuals to which a semantically relevant meaning can be attributed, depending on the respective application" [120]. Examples include a queue of people, the formation of uni-directional "lanes" in bi-directional pedestrian flows, the intersection of these lanes, or a group of people at a specific location. We use the phrase *texture of a crowd* to express the spatio-temporal relationships resulting from the interdependencies in the social fabric of a group of people. At this stage we may not yet fully understand what potential insights can be derived from crowd textures, but in any case it enables the study of emergent

The contents of this chapter have been originally published in **"Crowd textures as proximity graphs"** C. Martella, A. van Halteren, M. van Steen, J. Li, C. Conrado - *IEEE Communications Magazine* 2014, and have been slightly modified to improve readability.

spatio-temporal and social behavior of people in a crowd. For example, it allows one to question to what extent a group is dispersed in a crowd.

Discovering and investigating the texture of a crowd is at the heart of research in crowd management. As a prerequisite, it is essential to adequately represent texture. A common approach is to simply place cameras and collect their images and videos. There are a few drawbacks: the computational cost of video analysis limits the scale at which experiments can be run, cameras can be affected by complications such as occlusion and incomplete coverage, and privacy issues can emerge when the footage is recorded from real-world events.

As an alternative, on-body sensors can be used. Such sensors can collect rich information about the individual behavior of each subject. We believe a better understanding of crowd dynamics can be achieved by sensing from *within* the crowd instead of from an external observation point, as the sensing is based directly on the crowd: the individuals forming the crowd. An example of this approach can be found in [120] where accelerometers are used to recognize groups of people walking together.

A crowd is more than just a sum of individuals and collective behavior results from a continuous interaction and mutual influence between each individual and those nearby. The literature presents a vast number of examples of such behaviors as exhibited by animals such as swarms of insects, flocks of birds and schools of fish, and there is evidence of herding behavior also in humans [49]. Such networks of influence are fundamental for the emergence of collective behavior and are based both on the spatial relationship between the individuals as well as their social relationships.

In this chapter, we address the representation of the texture of a crowd through a *(dynamic) proximity graph*[1]. The proximity graph provides a computational representation of a crowd over time, it allows the analysis of the crowd texture it represents, and it provides a way to compute the effects of interventions into the crowd.

Modeling relationships between individuals through a graph is not new. Social graphs represent social relationships between individuals through edges. Also, on-body sensors have been used to actually measure social graphs [32, 74]. Finally, since a few years various groups have been dedicated to gathering data on the mobility of people.

[1] Our definition of proximity graph should not be confused with that of a Relative Neighborhood Graph, although the two share some of their properties.

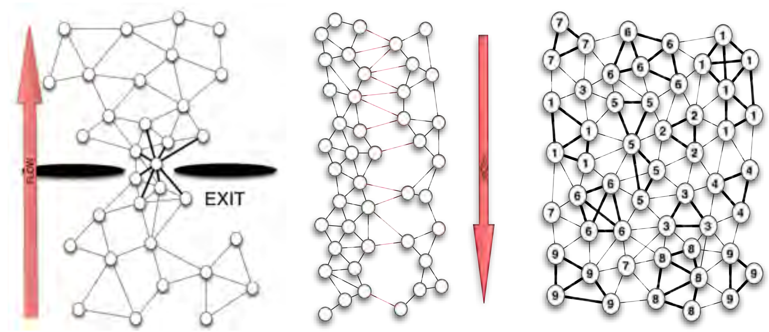
However, using a spatio-temporal graph to represent the texture of a crowd has not been done before, and we are not aware of any attempts to do so at the scale of (tens of) thousands of people. Besides its intended scale, the novelty of the proposed approach lays in the content of the proximity graph and the semantic interpretation of an edge. At a specific moment in time, an edge merely represents that two individuals were close to each other. Measurements over prolonged periods will reveal social groups (as we will discuss in this chapter), spatial structures (like lanes, clogging, and so on), but also the changes in the texture that result from targeted interventions (like displaying announcements on a large public display). The main contribution of this chapter is introducing the concept of crowd textures and their representation with proximity graphs.

The remainder of this chapter is organized as follows. We connect the local spatio-temporal nature of crowds and crowd dynamics to the concept of crowd texture. We then introduce the proximity graph as a representation of crowd textures, along with analytic examples. We describe a system architecture for an instrument to extract a series of proximity graphs from a crowd, analyze it, and communicate feedback to the crowd. We present an analysis of the proximity graphs we collected during a real-world experiment through a wearable device. We conclude by discussing possible extensions to the presented work.

3.2 Crowds and crowd dynamics

A generally accepted definition of a crowd is that it is a sizable number of individuals gathered together at a specific location with a sufficient density distribution, for a measurable amount of time and for a specific purpose. Moreover, the individuals in a crowd generally act in a coherent manner, sharing a social identity and common goals, interests and behavior, in spite of coming together in a typically unfamiliar situation.

Crowd behavioral patterns emerge from individual human interactions, which typically have a strong local character: individuals in a crowd influence one another, and this influence is stronger between nearby individuals. In fact, several models of crowd dynamics - which give rise to large-scale emergent behavioral patterns - do take into account local interactions between individuals, as well as the related measures of crowd density. For example, the social force model and its extension for panicking pedestrians [66] considers physical forces between individuals (as well as forces between individuals and the environment). The forces



(a) Congestion. (b) Flows. (c) Social groups.

Fig. 3.1 Examples of analyses to perform on the proximity graph.

depend on the dynamic spatial relationships among close individuals (more specifically, their distance). This model can be simulated to show most of the coordinated and uncoordinated behavioral patterns mentioned above.

From these continuous interactions, reflected in the crowd texture, behavioral patterns in crowd dynamics can clearly emerge. Therefore, by *measuring* the crowd texture, substantial information about the underlying crowd dynamics can be gathered. However, in order to be useful, this information needs to be put in the context in which the crowd exists. The contexts can be diverse, e.g., a football stadium, a train station, a music festival or a scientific conference. A pattern such as clogging, for instance, may occur in all these diverse contexts as it depends on environmental constraints such as the existence of narrow passages. In a music festival or stadium, clogging may occur at the entrance or exit of the festival site or stadium. Finally, in a train station, clogging may occur at the entrance to the station hall or entrance to a train, or at the access to stairs/escalators. The same measured crowd texture will clearly mean different things in different situations, with corresponding different risk level assessments and different feedback strategies to intervene in the crowd.

3.3 The proximity graph

A static proximity graph is a representation of the texture of a crowd at a specific moment in time. In its basic form, each vertex corresponds to an individual. Two vertices are joined by an edge if the two individuals they stand for happened to be in physical proximity, within a chosen distance. Each edge represents only the Boolean relationship without capturing the actual distance. Time is generally discretized into small slots. If and only if two individuals were *detected* to be in each other's proximity during a time slot, their associated vertices will be joined by an edge for that slot. A proximity graph is therefore *dynamic*; we speak of a proximity graph at time t as a *(proximity graph) snapshot* at t. Note that a proximity graph is naturally represented by a time series of proximity-graph snapshots. Although a proximity graph contains spatial information, it does not rely on absolute positioning data. It is the global representation of the texture of a crowd constructed from the local perspective of individuals within that crowd.

A proximity graph incorporates information to support the modeling of crowd dynamics where the individual behavior is defined only on relative neighborhood information. Basic models of flocking behavior are of this kind, as they are controlled by the following simple rules [119]. (1) Separation: avoid crowding neighbors (short range repulsion), (2) Alignment: steer towards average heading of neighbors, (3) Cohesion: steer towards average position of neighbors (short range attraction) . Each of the three rules can be applied by a member of the crowd based only on local information, namely its neighbors' states, and no global view or absolute spatial reference points are necessary. This type of modeling has been successfully applied to human herding behavior as well [49]. The above set of rules has been extended in different ways since its introduction to incorporate emotions, leadership etc.

3.3.1 Analyzing the proximity graphs

We now introduce a few examples of categories of analyses that could be performed on proximity graphs, which correspond to the contexts mentioned in the previous section.

Figure 3.1(a) presents a scenario where a crowd is exiting a stadium. As each individual approaches the exit, the local density increases given the physical bottleneck imposed by the gate. In the proximity graph, the local density is represented for each vertex by its degree, the number of incident edges (depicted in the figure for a particular vertex with thicker lines). Recognizing a congestion requires measuring the gradient of the average density of a crowd as it approaches the exit.

Figure 3.1(b) presents a scenario at a train station where a pedestrian flow is forming on a platform. There are two groups of individuals: on the left there is a stationary group waiting for a train, on the right there is another group that has just stepped out of another train and is heading towards the exit. The detection of the pedestrian groups is based on the transience of edges. In a time interval, the transience of edges, and their repetition, determines the degree to which neighborhoods remain the same. Intra-group edges present a relative stability. Instead, inter-group edges, depicted in red in the figure, are characterized by a short-living nature. By filtering out these edges, it is possible to detect the two connected components representing the two groups.

Figure 3.1(c) presents a scenario at a music festival. This type of event is usually attended by groups of socially related individuals that tend to stick together during its course. Nonetheless, groups might occasionally split e.g., to reach the bar, the rest room, or due to the density in front of the stage. In the figure, we picture a moment where a crowd stands in front of a stage. The edges represent the current proximity between the individuals, and the thickness represents the accumulated time spent close to each other over the whole period. Edges representing past proximity have not been drawn. In fact, they are not valid to describe the current crowd texture, but are used to compute the accumulated time. The detection of social groups requires preserving only the edges that present a long-lasting time interval when compared to the others. Each vertex is annotated with a group label it has been assigned to. Note that vertices that are far apart can still belong to the same group as some of its members might have currently and temporarily split.

3.3.2 The awareness of the context

Up to this point, we have considered vertices corresponding only to individuals. However, the proximity graph can be extended to represent other types of objects as well, e.g., a door, an ATM, or a food stand. This type of extension allows for a semantic and contextual interpretation of the observed behavior. In fact, different interpretations can be applied to the same crowd texture, depending on the context where the observation took place. Figure 3.2 presents an example of how the same proximity graph can be interpreted differently. Knowing it was collected in front of an ATM, the left-most graph can be interpreted as a queue. Instead, the same graph can represent an orchestra, knowing that the context was a stage (in this case the right-most vertex in the graph corresponds to the conductor while the other vertices correspond to the musicians).

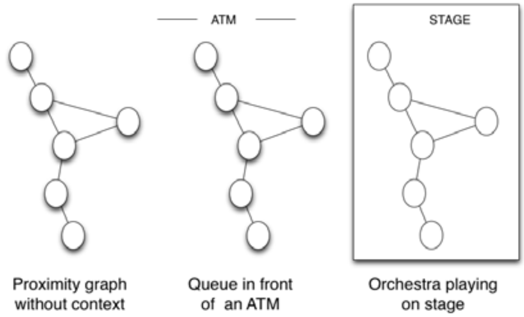


Fig. 3.2 Example of the dependency of the proximity graph on contextual information.

Location awareness allows for the disambiguation of the meaning a proximity graph can assume.

The concept of awareness can be generalized through annotations. An annotation to a vertex is a key-value pair that describes a specific property of the object represented by that vertex, e.g., gender, age, exit number, gate capacity, GPS coordinate, staff role, etc. Annotations contribute to awareness as they allow to picture more precisely the observed texture and support informed decision-making. For example, through an annotated graph it is possible to recognize at what specific exit a congestion has formed, if children are involved, or whether stewards are already around the area for support. Also, static vertices with GPS coordinates annotations can act as absolute positioning reference points when necessary.

A consequence of the relative nature of the proximity graph is that movement cannot be attributed to vertices. An edge "breaks" when at least one of the two individuals involved in the relationship moves away. As the result of these actions is the same, determining which one occurred is not possible. Going back to the platform example in Figure 3.1(b), recognizing the short-living edges, depicted in red, allows for the detection of relative movement between the two groups. Again, it is not possible to determine for each group whether it is moving or not. Extending the graph with static vertices spread along the platform provides a solution to the problem. In fact, the vertices belonging to the moving groups would create short-living edges with these static vertices allowing inference of movement.

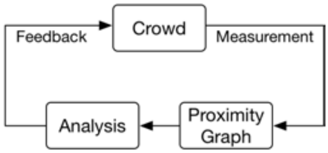
The temporal information encoded in the proximity graph allows for more than the description of crowd dynamics. As proximity can be a by-product of social interaction, socially related individuals tend to spend more time close to each other than strangers. This principle is at the basis of the example depicted in Figure 3.1(c). For example, in [32] the authors were able to analyze the social ties within organizations by looking at low-frequency proximity information. In general, knowing how much time two or more individuals have spent close to each other, with possible addition of contextual information about location, enables the inference of the type of social relation incurring between them.

Temporal information also enables the inference of consequentiality. Imagine the festival scenario depicted in Figure 3.1(c). Suppose we observe individual A conversing first with B and later with C. If, later on, we would note A, B and C conversing together, we could imagine that B and C were introduced by A. In the same way, we could predict that an individual who was observed in proximity to a counter, where tickets for drinks are sold, will eventually show up at the bar. Similarly, the temporal dimension of face-to-face interactions has also been investigated to study the spreading patterns of epidemic diseases in different types of social events, e.g., a conference or a museum [74].

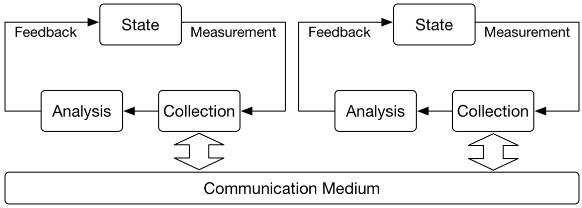
3.4 An instrument

We identify the following requirements for an instrument to measure and analyze the proximity graphs presented thus far. (1) It is composed of a device that is *wearable* by an individual, and (2) that is *detectable*, within a chosen distance range, to measure proximity. (3) The information measured by the devices is *extracted* and *collected* through an infrastructure. (4) The collected proximity graph is *analyzed* by a processing system looking for patterns, and (5) *feedback* is transmitted from the system to the individuals.

One notable way to implement our sensing device is an application running on a mobile phone. Mobile phones are widely diffused across the population, and already ship with both high computing power, e.g., dual-core processors, and a broad range of sensors, e.g., accelerometers, microphones, cameras etc. On the other hand, with the decrease in price and size of wearable technology, another way of implementing our device is through a smart chip-card, such as those currently used for public transport tickets or for badges on the working places (the same technology can be easily integrated into a festival bracelet as well). Figure 3.5(a) shows the device we currently use for our experiments with the proximity graph. We return to its description shortly.



(a) Centralized processing chain.



(b) Decentralized processing chain.

Fig. 3.3 Block diagrams for the processing chain.

3.4.1 Sensing technologies

As far as detectability is concerned, multiple technologies have been investigated in the past. Infrared, ultrasound and radio-frequency sensors, e.g., RFIDs and Bluetooth, have been utilized to track face-to-face interactions and co-location. The general approach to detection requires the assignment of a unique identifier to each device. The ID is periodically transmitted to nearby devices over the communication medium, and the reception of such a message constitutes a detection.

Various aspects influence the functionality of proximity detection, and determine the vertices' neighborhoods. The range, direction, and angle of the transmission cone bias the type of interaction being recorded. For example, face-to-face interaction is tracked through a transmission range within about 2 and 4 meters, a frontal direction, and a narrow cone of about 20 degrees, e.g., via an infrared sensor. Conversely, co-location is measured through long-range omnidirectional transmission, e.g., via Bluetooth. The theory of Proxemics guides the choice of transmission range depending on type of social behavior that one wants to measure. Measuring a crowd texture poses a set of constraints to the detection strategy: omnidirectionality to maximize the recall of nearby devices, short-range transmission to detect only close-by devices, and high-frequency to grasp instantaneous changes to the texture.

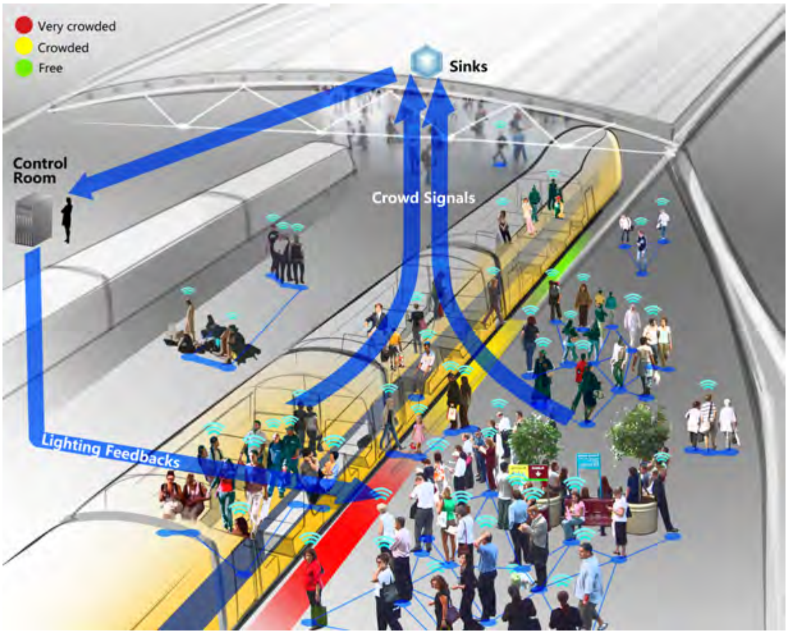


Fig. 3.4 How to sense and use the texture of a crowd at a train station. The density inside the train and on the platform is used to guide the individuals towards the less crowded coaches.

3.4.2 System architecture

As neighborhood information needs to be extracted and collected from the devices, they need to be connected to a network that allows them to reach a central repository. The mobility and the high density, which characterize a crowded environment, make centralized networks, such as cellular networks and WiFi, unsuitable for this utilization. In contrast, decentralized ad hoc wireless networks provide the flexibility to design problem-specific protocols that guarantee higher scalability. Typically, such networks make use of special devices, usually called *sinks*, that receive data from on-body devices to subsequently store that data at a central repository. Sinks are spread around the event location to achieve high coverage. The on-body devices can reach the sinks by using a high transmission range, or through dissemination protocols, e.g., gossiping and routing. The sinks bridge the ad hoc wireless network with the network where the central repository and the processing systems are situated. The central repository and the processing system can be deployed to one or multiple servers at the event location or in a cloud service.

We connect the devices through an ad hoc wireless network, as it matches the given constraints while enabling both the transmission of detection messages and the exchange of information between the devices. Also, it allows for the communication with the external network for processing and feedback.

While data streams from the devices to the central repository, the global view of the proximity graph gets constantly updated. Periodically, the processing system analyses the proximity graph, examining the crowd texture for known patterns. Recognition of crowd dynamics is a classification problem that requires statistical analysis of the metrics of the graph, either at vertex, group, or global level. The temporal information in the proximity graph is exploited by analyzing snapshots, defined by time intervals whose duration depends on the analysis being performed. Graph algorithms are potentially computationally expensive, introducing additional latency between the measurement and the recognition of patterns. This requires a system with sufficient computational power to analyze a proximity graph within the required time constraints.

Once the information is extracted from the proximity graph, it can be presented for feedback. The feedback can be sent to the crowd managers or directly to the crowd, either through the devices or through fixed infrastructure, e.g., screens, speakers, etc. While feedback can reach the control room and the infrastructure over the reliable network, devices can be reached via the ad hoc wireless network. In the same way information is extracted and collected from the devices, dissemination protocols allow feedback to reach specific individuals, groups or the whole crowd. The intervention strategy defines the information, the destinations, and the techniques used to give feedback to the crowd.

3.4.3 A processing chain

We present now a processing chain that incorporates all the components presented above into a loop. The processing loop is presented in Figure 3.3(a). Each individual wears a device with a proximity sensor capable of detecting the other devices within a chosen distance range. The devices share a communication medium that enables transfer of information. The loop starts with the measurement of the individual's neighborhood through the device's sensor. Once the neighborhoods are measured, the next step consists in the collection of the neighborhoods from the devices to a central repository, to compose the global view of the proximity graph. Afterwards, the proximity graph can be analyzed by a processing system to recognize crowd dynamics. Possibly, feedback is computed, and sent to the managers or to the crowd. At this point, the new state of the crowd can be measured and the loop can start again.

Figure 3.4 shows a possible future instantiation. People at a train station are assumed to have proximity sensors, for example, embedded into their smart phones, allowing for the detection of a proximity graph. The analysis of the situation in the train and on the platform may be used to subsequently inform people where to board, or to leave the train.

Note that centralization is not strictly necessary to the instrument. As individual behavior depends on local context represented by the individuals and the objects in close proximity, the analysis algorithm can often be expressed based on a vertex-centric local view of the graph. Following this approach, nearby devices can exchange their neighborhoods right after the measurement step, and analyze their local view of the graph autonomously, without relying on a global view of the graph contained in the central repository. This autonomy decreases the interval between the moment the state of the crowd is measured, and the moment feedback can be generated. This alternative is presented in Figure 3.3(b). However, a third possibility is available: a hybrid system where the devices locally aggregate and process proximity information, and only these aggregated views are later collected and processed centrally. This latter approach provides both decreased latency in feedback generation and central monitoring of the crowd. Moreover, it minimizes the overhead of data extraction from the devices, as only aggregated information is transmitted.

3.5 Real-world experiment

To give a flavor of what can be achieved following our approach, in this section we describe an example application based on a real-world experiment we conducted at an ICT conference. The conference was divided into 7 tracks, each focusing on a specific ICT topic, such as High-Performance Computing, Software Engineering, Security, etc. Of the 250 attending individuals, 139 were wearing one of our devices as a name tag throughout the whole day. We asked each of the participants for their main track of interest, as a hint to community membership (note that the groups were not balanced, as for example one had only four individuals participating in the experiment). Table 3.1 shows the distribution of the participants across the tracks. The device, depicted in Figure 3.5(a), has an Atmel ATXMega128 CPU with 8k of RAM, 128k of flash memory, and a Nordic nRF24L01+ wireless radio. To

Track	1	2	3	4	5	6	7
Participants	27	9	19	4	26	10	27
In LCC	15	7	13	3	20	4	22
Correct Classifications	0.53	1.0	0.62	0.0	0.74	0.0	0.77

Tab. 3.1 Sample statistics and results of the analysis. Additional 17 participants were part of the organization and were not labeled with one of the main tracks.

communicate, the devices create an ad hoc wireless network through an energy-efficient MAC protocol designed for mobile social networks [44]. Through this network, every second each device transmits its ID to the devices nearby, within some 2-3 meters of distance, allowing for its detection.

The devices log detections on the on-board storage unit along with their timestamps. At the end of the event we downloaded these logs from the devices for offline analyses. The devices were also broadcasting a second type of transmission. Between two short-range transmissions, a long-range transmission was broadcast up to about 20 meters. This second broadcast contained the list of IDs received by each device during the previous second, hence comprising only short-range detections. We captured these long-range transmissions through the "sinks" we installed in the main hall of the event location. We used this data to visualize at the event the evolution of the proximity graphs in real time. Long-range transmissions were not used to detect proximity.

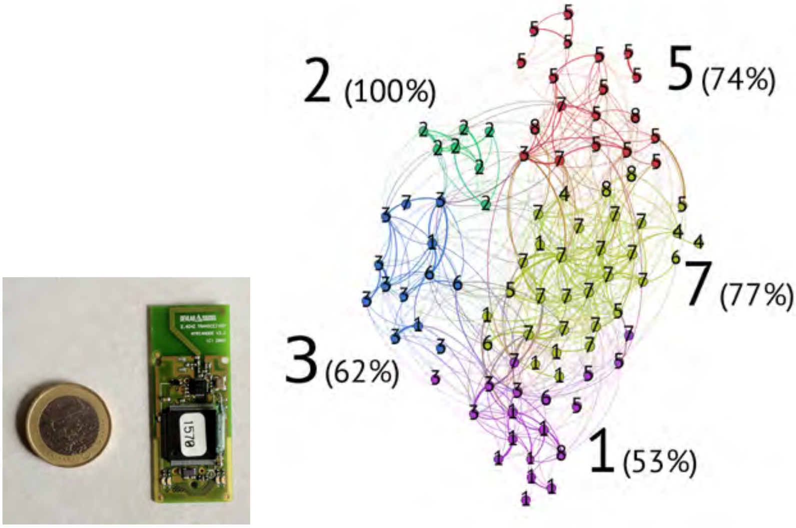
Although we recorded proximity information during the whole conference, we concentrate here on the two hours between 12:00 and 14:00, when the poster session and the lunch break took place. During this time, all the participants gathered in the main hall. Our goal was to investigate to what extent during this time the participants stayed close to people they share interests with, as indicated by their main track of interest. To perform this analysis, we aggregated the series of proximity-graph snapshots into a single undirected static graph, for which we decided to join two vertices by an edge if the two corresponding individuals had been in physical proximity for at least 600 seconds during the two hours. Each edge has a weight that accounts for the total number of seconds the two individuals have spent in physical proximity. On the largest connected component (LCC) of this graph, we ran a state-of-the-art community-detection algorithm [116].

The analysis we performed is analogous to the detection of social groups presented earlier. The community-detection algorithm assigns vertices to communities trying to maximize modularity. Graphs with high modularity tend to have dense connections between vertices within communities and sparse connections between vertices across different communities. Intuitively, it groups together vertices that are interconnected and have spent long time together. Figure 3.5(b) shows the results of the analysis. Vertices are colored according to the community they have been assigned to by the algorithm, and labeled according to the main topic of interest. The clustering tends to assign vertices with the same label to the same community, supporting our hypothesis and showing the validity of the data extracted through the instrument. Note that the algorithm does not make use of the noted information about interests, but only of the topology of the graph. One should not consider the main track of interest as ground truth. In fact, many of the individuals indicated their interest as one of out of more possible ones. Moreover, as the participants came from a number of universities and departments, they tended to socialize also according to different criteria, for example, with people with whom they shared affiliation. Finally, the nature of poster sessions and banquets stimulate people to spend time in proximity to people belonging to different communities. We leave a deeper and more sophisticated analysis of the experiment for future work.

The unreliability of the wireless communication and of the devices causes mis-detections, meaning that detections are missed or erroneously added. Typical examples of these causes are "collisions," interferences of other sources of radio-frequencies (such as WiFi spots), the shielding of the human body, data corruption due to faults in the device, etc. For this reason, to extract the proximity graphs used in our analysis we pre-processed the logs with a density-based clustering algorithm. The algorithms exploits the "bursty" nature of the collected data to reconstruct part of the missed detections and filter out noise. Our results show that this filtering phase greatly increases the sensitivity of the instrument [98] (see Chapter 4).

3.6 Discussion

The local nature of individual behavior in crowds also affects the analysis of crowd dynamics through other modalities. In [120], the authors propose a processing chain for the recognition of crowd behavior from mobile sensors with pattern analysis and graph clustering. Subjects wear on-body sensors, and move collectively. Activity patterns are then extracted from the sensors for individual behavior recognition, e.g.,



(a) Experimental device. (b) Community detection at a conference.

Fig. 3.5 (a) Experimental device. (b) Colors indicate the detected communities, labels indicate the main topic of interest of the individuals, edge thickness indicates the total amount of time the individuals have spent in physical proximity. The vertices have been positioned through a spring-embedding algorithm. The algorithm takes into account only the topology of the graph, hence the layout is not related to location information within the main hall. We have labeled each community with the label that appears most frequent in it, and the value expresses the percentage of vertices with that label appearing in their community.

from accelerometer data. Afterwards, a pairwise correlation of this information between each pair of individuals is computed, forming what they call a disparity matrix. Finally, this information is transformed into a graph by performing multidimensional scaling. In the obtained graph, each vertex corresponds to a subject. The neighboring nodes have similar behavioral characteristics, and are thus more likely to participate in the same dynamics. By performing a graph clustering algorithm, they are able to predict that the subjects corresponding to the vertices belonging to the same cluster participated in the same group. By exploiting the information contained in the proximity graph, only the behavioral data between neighboring vertices could be compared, reducing the cost of an expensive step of the processing chain. The proximity graph is a representation of the crowd texture that can be used either to directly recognize crowd dynamics, or to support recognition through other modalities.

The mutual influence between individuals is an interesting aspect of social behavior, as it can be used to guide a crowd by targeting a subset of the individuals through feedback information about the current state. Although during the last few years, we are starting to better understand crowd dynamics, less is known about how to influence a crowd. The proximity graph is a representation of a crowd and it allows to compute interventions towards a desired behavior. Consider the following (admittedly still speculative) examples of simple interventions on crowds as presented previously in their scenarios.

The example in Figure 3.1(a) could represent a bottleneck at the entrance of a stadium. This is a typical situation where people get pushed, and in the most dramatic conditions also walked over. This behavior often finds its origins in the absence of information. The individuals at the back of the crowd cannot see the high density at the entrance, or the presence of a congestion, and may start to push. A screen on top of the gate, visible for all the individuals, could depict with colors the density in the front.

The example in Figure 3.1(b) could represent a platform in a train station. A common situation in such a scenario is that people getting out of the train tend to head towards the closest exit. Uneven usage of exits could be avoided by feeding back information about less crowded exits.

Finally, the example in Figure 3.1(c) could represent groups of visitors to a festival. Such events are usually visited by groups of friends. It is common that people lose contact with some members of their group. Once the groups are detected, each member of the same group can be guided towards the same exit, so that they can find each other there.

These are just a few examples of the many possibilities that emerge once the texture of a crowd is captured in a proximity graph. Our approach has great potential to accelerate the emerging field of Computational Social Science. In particular, the capability of sensing the crowd from within, without any requirement for location information or centralization, in respect of the privacy of the individuals in the crowd. Moreover, it allows to compute timely insights about the state of the crowd, and communicate feedback to ensure safety and comfort.

3.7 Conclusions

We have presented the concept of a crowd texture as the set of spatio-temporal relationships between the crowd members that results from their behavior. We have presented a series of proximity graphs as a representation of such crowd textures. We have characterized the properties and limitations of the model, together with its extension to person-to-object relationships, and additional information about the context in which the crowd is embedded.

A number of approaches were suggested to identify a number of selected crowd dynamics, as algorithms that mine a series of proximity graphs. We have described a distributed architecture of a system to measure, collect, and store the series of proximity graph, as well as process the graphs to identify crowd behavior. Finally, we have presented a real-world use case where we have applied a graph-clustering algorithm to a series of proximity graphs collected during an ICT conference, to demonstrate that social groups in the crowd could be detected.

The approach is easily applicable to many problems in the context of crowd and crowd management, and it is general to the medium and architecture used to collect the spatio-temporal information about the crowd, as long as it models after relative-proximity information.

4 From proximity sensing to spatio-temporal social graphs

> *Failure is not an option here. If things are not failing, you are not innovating enough.*
>
> — **Elon Musk**

4.1 Introduction

In the previous chapter we have discussed how a crowd texture can be captured with proximity sensors, collected to a central repository via a distributed system architecture, and finally analyzed to study group behavior. We have also presented a case study where a graph clustering algorithm was used to identify social groups. In that work, we first applied a filtering algorithm to overcome the limitations of noisy radio-based proximity sensing that cause loss of proximity detections and the collection of spurious ones. In this chapter we present the details of the filtering algorithm used in the mentioned case study to re-construct a series of proximity graphs from a collection of proximity detections.

While proximity sensing has been used to study face-to-face interactions in a variety of contexts and applications, from the relationships in the work place [110, 159, 43], to co-location [6, 50], to the dynamics in a crowd [74, 79], these techniques rarely deal with the specific characteristics of the device, the medium, and the environment, which cause noisy and lossy measurements. They are designed to be deployed to more controlled environments (e.g., an office, a lab, a conference hall) than those where usually crowded events take place. In this chapter, we

The contents of this chapter have been originally published in **"From proximity sensing to spatio-temporal social graphs"** C. Martella, M. Dobson, A. van Halteren, M. van Steen - *IEEE International Conference on Pervasive Computing and Communications PERCOM* 2014, and have been slightly modified to improve readability.

describe and evaluate a method that provides a reliable, fine-grained, and generally applicable proximity measurement.

Challenges. Designing a method to sense proximity poses a variety of challenges at different levels. At the *technological* level, the sensor must detect, in absence of physical contact, physical proximity between individuals within a specific range. Typical examples are infrared, ultrasonic, and radio-frequency sensors. This measurement should be performed consistently within a well-defined range, and as uniformly as possible over this delineated area. As the device might be worn for an extensive amount of time, it should be unobtrusive, ergonomic, and energy-efficient.

At the *processing* level, the collected detections must be filtered to overcome the inherent limitations of the technology, the medium, and the environment. These limitations provoke lossy and noisy measurements, meaning that detections might be missed or erroneously added, for example due to corruption or interference to the signal. The ability of the method to cope with these disturbances is a strong requirement. In fact, the quality of the measurement will be defined by the extent to which the method will be able to deal with these limitations.

In this chapter, we propose a method to reliably measure physical proximity based on a device composed of a sensor that can detect other sensors nearby and a processing unit that can filter these detections, and a filtering algorithm. In our experiments, we used an ad-hoc device which allowed us to deal with the challenges at both levels. However, our approach is applicable to other devices as well, such as mobile phones. We dealt with the challenges that belong to the technological level in [45, 44]. In this chapter we focus mostly on those that belong to the processing level.

Our approach. We assume that each individual wears a device with an associated unique identifier. Periodically, each sensor broadcasts its identifier, and this transmission is received by the sensors within a range d. We say that sensor u *detects* sensor v at time t if it receives that particular transmission containing the identifier of v. Conceptually, each sensor collects these detections along with the time of reception t. If we model every device as a vertex, we obtain a series of what we call *proximity graphs*. Each proximity graph in this series represents a unit of time, and it contains an edge from vertex u to vertex v if sensor u detected sensor v at that specific time. In principle, the proximity graph

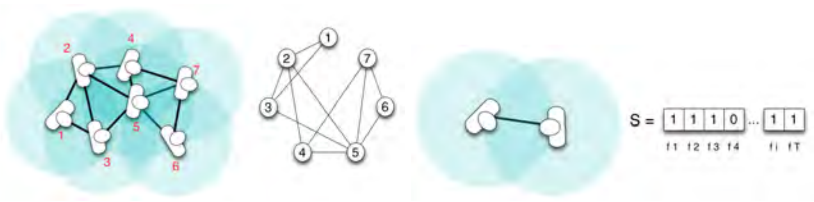


Fig. 4.1 (Left) A proximity graph representing a group of individuals. The blue disc represents the detection range. (Right) A series of detections between two sensors represented through a bit string.

is a directed graph, but we can drop the edge direction and consider the detection as symmetric. Figure 4.1 pictures an example of a proximity graph.

The proximity graph is a representation of proximity between individuals, and as such it can represent the *texture* of a crowd [97] (see Chapter 3). It can be used to study the underlying social behavior. Note that the proximity graph is a type of spatio-temporal graph, but it does not contain any location information, such as the absolute position of the individuals, or information about distance or angle of detection. An edge is "only" a boolean relationship, as it connects two individuals and it represents whether or not they were close enough to each other at a specific time. Although minimal, the evolving series of proximity graphs contains information that can be used to study systems such as epidemic models, crowd dynamics as queues and pedestrian lanes, social ties, etc. Moreover, it can be augmented with other sources of information, such as profiles or different sensor data, to correlate patterns of proximity with various aspects of social behavior.

We want to construct the series of proximity graphs over time through a number of devices. In particular, we want to perform the measurement reliably, filtering out the noisy detections and reconstructing the missing ones. To this end, we follow a *data-mining* approach where we focus on higher-level analyses of the collected data instead of trying to solve the problem at a lower layer, such as by improving the sensor or the MAC protocol. This work makes the following primary research contributions. We present the design of a method based on a radio, a protocol used to create a social ad-hoc wireless network, and a processing pipeline to improve the quality of the proximity graph by reconstructing missed detections and eliminating false detections. We evaluate the reliability of the approach both in simulation and with real-world experiments.

4.2 Model

In this section we present the problem definition and the proposed solution, but first we proceed by introducing our model. We consider a time interval which we divide into T *frames*, each having a duration of γ time units. In this chapter, γ is taken to be 250, 500, 750, or 1000 ms. Consider a series of detections $S(u,v) = \langle b_1, \ldots, b_T \rangle$ between two sensors u and v. In this case, $b_i = 1$ if and only if a detection message sent by v was successfully received by u in frame i. Hence, S can be practically represented through a bit string. Theoretically, more than one transmission can take place, but we assume that detection messages are sent regularly with a frequency of $\frac{1}{T}$. We will informally speak about a detection at time i, to mean a detection during frame b_i, and indicate frame i with bit b_i. We aggregate the detections of u about v and vice versa together in a single string, by performing a logical *OR* between the respective bit strings. This design decision was taken to overcome missing data, but a logical *AND* could also be possible, for example in case a stronger requirement of symmetry is necessary. Figure 4.1 shows an example of a bit string representing the proximity detections between two sensors.

4.2.1 Problem definition

Ideally, if two sensors are within a given threshold d of physical distance for an interval of time, they should be able to detect each other at every time. This would produce an ideal series S where subsequences of set bits are contiguous for the whole duration of the proximity. Extracting the edges of the proximity graphs from such a string would have a trivial solution. However, the incomplete and noisy measurement produces mis-detections that cause incorrect values to be assigned to some frames. The consequence of these incorrect bit values is a number of excessive or missing edges in the proximity graphs. To overcome this, a more sophisticated solution is required. Fortunately, the nature of the measurement is *bursty*. This means that correct detections tend to appear together, and false detections tend to be isolated. This nature is inherent due to the social behavior being measured, as individuals tend to stay close for a certain interval of time, and then separate [70]. We want to identify these bursts in S. Once bursts have been identified, we can can *set the unset bits* within the bursts, hence correcting the missed detections. Also, we can *unset the isolated bits* that are not part of any burst, hence eliminating false detections. Moreover, while minimizing

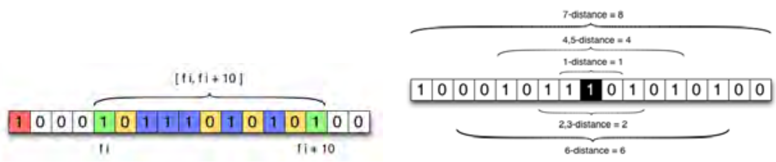


Fig. 4.2 (Left) A series of detections after applying density-based clustering. Blue frames are part of the core of the cluster, while green frames are density-reachable from them. The red frame represents noise and the bit in it will be unset, while the bits in the yellow frames will be set. (minPts=3, epsilon = 2) (Right) The k-distance of b_i for different values of k. The black frame indicates detection b_i.

the number of missed detections, we also want to minimize the number of false detections introduced by the method.

The bursty nature of the measurement implies that a detection is often surrounded by other detections within a certain window of time. This means that S is characterized by set bits surrounded by other set bits nearby. We can obtain a *smoothed* version S' of S by applying a window-based smoothing technique to S as follows. A window is defined as $window(S,i,n) = \{b_j \in S \mid j \geq i - \lfloor \frac{n}{2} \rfloor \wedge j \leq i + \lfloor \frac{n}{2} \rfloor\}$, with $n > 1$ and odd (we ignore the case where $n = 1$ as the technique has no effect). Parameter n defines the size of the window. Given a bit string $S = \langle b_1, \ldots, b_T \rangle$, we define S' as $S' = \langle b'_1, \ldots, b'_T \rangle$ where $b'_i = 1$ iif $\|\{b_j \in window(S,i,n) \mid b_j = 1\}\| > \rho$, with $\lfloor \frac{n}{2} \rfloor < i \leq T - \lfloor \frac{n}{2} \rfloor$. We define $b'_i = 0$, with $i \leq \lfloor \frac{n}{2} \rfloor \vee i > T - \lfloor \frac{n}{2} \rfloor$. Informally, we let the window *slide* over S to produce a new bit string where at each slot a bit is set if and only if there are enough set bits around the corresponding bit in the original string (including that bit). This effectively sets the unset bits in a burst and unsets the isolated ones (when $\rho > 1$). While simple and effective, the behavior of the technique depends on the choice of the two parameters n, ρ, for which no well-informed heuristics exists, hence limiting the applicability of the technique. In the next section, we propose an alternative technique that behaves similarly, but permits an estimation of its parameters. The strength of this approach lies in the consistency between the estimation technique and the semantics of the estimated parameters.

4.2.2 Density-based clustering

Given the bursty nature of S, we can use a density-based clustering algorithm and cluster dense subsequences of bits together. These clusters would correspond to the bursts of detections we are looking for, while the outliers would correspond to the noisy detections to be ignored. In particular, we propose to use DBSCAN [55]. In S, the distance between two bits b_i and b_j is defined as $dist(b_i, b_j) = |i - j|$. A bit b_i is *directly density-reachable* from another bit b_j if their distance is not more than ε (in that case it is said to be part of its ε-neighborhood) and if b_j is surrounded by a sufficient number of bits to consider them a cluster, denoted as $minPTS$. A bit b_i is called *density-reachable* from b_j if there is a sequence of bits $b_1, \ldots, b_k, \ldots, b_n$ with $b_1 = b_j$ and $b_n = b_i$ where each b_{k+1} is directly density-reachable from b_k. Two bits b_i and b_j are *density-connected* if there is a third bit b_k that is density-reachable from both. A cluster is a subset of S composed of mutually density-connected bits, plus any bit that is density-connected to them. A point that is not part of a cluster is considered noise. Figure 4.2 shows the result of the clustering algorithm on a series of detections between two sensors.

Analysis. One can see a correspondence between the window-based smoothing technique and the density-based clustering technique. In both cases, we identify bursts by thresholding the number of set bits within a certain "radius", and we set the unset bits in the burst. In the former, we set a bit in S' if there are at least ρ set bits within a certain distance, defined by the window size n, in the original string. In the latter, we set the unset bits between density-reachable bits, which are similarly identified based on the minimum number of set bits *minPTS* within an ε-neighborhood. For this reason, one would expect to obtain similar results through the two techniques. Note that the application of DBSCAN to our model affords a more efficient implementation than general DBSCAN. The runtime complexity of general DBSCAN is $O(N^2)$, as each data point needs to be evaluated against all other data points to find its ε-neighbors. This runtime complexity can be reduced to $O(NlogN)$ by means of an accelerated indexing structure such as a spatial index. In our model based on bit strings, collecting the ε-neighborhood of a bit has a constant cost as it requires to evaluate the bits within the ε-distance. As ε is fixed and evaluating a bit in a string has a constant cost, the cost of DBSCAN applied to our model is linear to the number of bits in the string. This makes the runtime cost of the density-based clustering technique comparable to the cost of the window-based smoothing technique.

DBSCAN does not need any training phase and it does not require to define a number of clusters in advance. However, it does require that the data corresponds to a single density distribution function. In DBSCAN, the expected density distribution is defined through the parameters ε and $minPTS$, and the algorithm is very sensitive to them. As we do not know the density distribution in advance, we need to estimate these parameters. We propose such a technique in the next section.

4.2.3 K-nearest neighbors analysis

To study the density distribution of S we propose to compute the distance within which a bit will find at least k bits, by also taking into account that more bits can be at the same distance. By combining this notion of distance with the number of bits within that distance, we can estimate density. In particular, we propose to compute the distance of the *k-Nearest Neighbors* (k-NN), denoted as k-distance(b_i). Given the definition of $dist(b_i, b_j)$, the k-distance of detection b_i is the δ such that:

(i) $\|\{b^* \mid b^* = 1 \wedge dist(b_i, b^*) \leq \delta\}\| \geq k$.

(ii) $\|\{b^* \mid b^* = 1 \wedge dist(b_i, b^*) < \delta\}\| < k$.

$\|\{b_j, \ldots, b_k\}\|$ denotes the cardinality of the set and hence the number of detections. Figure 4.2 shows the result of the k-NN analysis run on a series of bits.

Computing the k-distance of each detection in the string permits an overview over the relationship between distance and neighborhood size. These in turn relate to the two parameters of DBSCAN ε and $minPTS$. In fact, the k-distance of b_i is the ε necessary to collect *at least* k other detections around it. As such, it can be used to select the correct ε that satisfies $minPTS$ for the detections in the bursts. Usually, $minPTS$ is known in advance and is bounded to the definition of a cluster and the noise in the data, and it can be used as a basis for the value of k. With large enough clusters, the detections should present a common k-distance that corresponds to a representative density of the clusters, except for the noise or the members of clusters smaller than k. Selecting an ε that is close to this common k-distance guarantees that DBSCAN will cluster the detections together with the others within the k-distance.

The *k-distance plot*[1] is a means to study a k-distance distribution. This plot is constructed as follows. For a particular value of k, the k-distance for each point is computed. These values are sorted in ascending order. The sorted list of points is plotted on a Cartesian plane, by using the k-distance value as the y coordinate and the position in the sorted list as the x coordinate. We will use and show k-distance plots in the evaluation sections, to select the DBSCAN parameters. Note that in this work we selected ε manually to show the impact of the different choices on the behavior of the method. However, approaches have been proposed to select ε automatically through a numerical analysis of the k-distance plot [158].

4.2.4 Proposed solution

Our proposed solution to extract a series of proximity graphs from a series of detections consists of the following steps. The detections are collected with their timestamps from each sensor, and they are aggregated and converted to a bit string representation. The k-NN analysis is performed on all the bit strings together, by computing the k-distance of each detection in each string, and sorting all these values together into a single list. At this point, depending on the technology being used to perform the measurement, usually a range of valid values of $minPTS$ is known in advance and it is used as k to compute the k-distances. Finally, each bit string is clustered with the representative k-distance as ε and k as $minPTS$. The extracted clusters and noise are used to correct the mis-detections, and finally the series of proximity graphs can be generated.

4.3 Evaluation in simulation

In this section we present the evaluation in simulation of our proposed solution. The objective of this evaluation is to validate whether the method can construct correctly a proximity graph in a controlled environment, in the face of both data loss and noise.

4.3.1 Methodology

For a scenario, we selected a social environment where individuals walk around and form groups dynamically, to simulate a social location such

[1] In the literature this is usually called *k-distance graph*, but we use the term *plot* to avoid confusion.

Setup	Raw measurement Sensitivity	Raw measurement Specificity	Density-based Sensitivity	Density-based Specificity	Window-based Sensitivity	Window-based Specificity
100slots0SNR1000ms	0.972 ± 0.002	0.999838 ± 0.000008	0.973 ± 0.002	0.99970 ± 0.00003	0.982 ± 0.001	0.99874 ± 0.00004
100slots0SNR750ms	0.972 ± 0.002	0.999876 ± 0.000008	0.972 ± 0.002	0.99983 ± 0.00001	0.982 ± 0.001	0.99894 ± 0.00004
100slots0SNR500ms	0.972 ± 0.002	0.999926 ± 0.000003	0.973 ± 0.002	0.999915 ± 0.000003	0.982 ± 0.001	0.99926 ± 0.00003
100slots0SNR250ms	0.973 ± 0.002	0.999986 ± 0.000001	0.973 ± 0.002	0.999986 ± 0.000000	0.978 ± 0.001	0.99979 ± 0.00001
100slots10SNR1000ms	0.856 ± 0.008	0.999997 ± 0.000001	0.857 ± 0.008	0.999989 ± 0.000001	0.876 ± 0.007	0.999949 ± 0.000004
100slots10SNR750ms	0.856 ± 0.008	0.999998 ± 0.000001	0.857 ± 0.008	0.999997 ± 0.000001	0.872 ± 0.007	0.999985 ± 0.000002
100slots10SNR500ms	0.857 ± 0.008	0.999999 ± 0.000001	0.857 ± 0.008	0.999999 ± 0.000001	0.867 ± 0.008	0.999995 ± 0.000001
100slots10SNR250ms	0.857 ± 0.008	0.999999 ± 0.000000	0.857 ± 0.008	0.999999 ± 0.000000	0.862 ± 0.008	0.999999 ± 0.000000
4slots0SNR1000ms	0.54 ± 0.01	0.996899 ± 0.000001	0.938 ± 0.006	0.99746 ± 0.00007	0.932 ± 0.006	0.9971 ± 0.0001
4slots0SNR750ms	0.53 ± 0.01	0.999997 ± 0.000000	0.946 ± 0.003	0.99941 ± 0.00006	0.927 ± 0.006	0.9974 ± 0.0001
4slots0SNR500ms	0.53 ± 0.01	0.999998 ± 0.000000	0.95 ± 0.01	0.99975 ± 0.00003	0.921 ± 0.008	0.9979 ± 0.0001
4slots0SNR250ms	0.51 ± 0.02	0.999999 ± 0.000001	0.93 ± 0.01	0.99995 ± 0.00001	0.892 ± 0.016	0.9988 ± 0.0001
4slots10SNR1000ms	0.33 ± 0.01	0.999999 ± 0.000000	0.79 ± 0.02	0.99961 ± 0.00008	0.799 ± 0.009	0.9987 ± 0.0001
4slots10SNR750ms	0.33 ± 0.01	0.999999 ± 0.000000	0.797 ± 0.008	0.99961 ± 0.00007	0.79 ± 0.01	0.99923 ± 0.00005
4slots10SNR500ms	0.32 ± 0.01	0.999999 ± 0.000000	0.809 ± 0.008	0.99988 ± 0.00003	0.778 ± 0.009	0.99971 ± 0.00002
4slots10SNR250ms	0.31 ± 0.01	0.999999 ± 0.000001	0.79 ± 0.01	0.999987 ± 0.000009	0.74 ± 0.01	0.99997 ± 0.00001

Tab. 4.1 Comparison of the results of the reconstruction of the proximity graph from raw data and with the two techniques. The results present the average outcome over the 10 mobility patterns for each combination. For the density-based clustering technique, the chosen parameters were $minPTS = 1$, and $\varepsilon = 5, 8, 8, 14$ for each group of scenario respectively. The values for ε were chosen by studying the k-distance plot for each group. For the window-based smoothing technique, the chosen parameters were $\rho = 1$, and $n = 3, 5, 7, 9$ for each group of scenario respectively. The values for n were chosen experimentally to obtain comparable results with respect to both sensitivity and specificity.

as a museum or a conference. Here we describe the different components we used to simulate our scenario and the metrics we used to evaluate the performance of our solution.

Mobility patterns. We generated 10 mobility patterns with BonnMotion [8] using the Reference Point Group Mobility (RPGM) model [68] for a simulated crowd of 100 individuals moving for 60 minutes inside of a square arena of 25x25m. Each pattern was generated through a different random seed, but with the same set of parameters. The generated patterns lead to a list of (x, y) coordinates describing where each individual will be at any given time. The individuals move with a speed of between 0.3 and 1.2 meters per second, but they can pause for up to 60 seconds. They form groups of a minimum size of 2 individuals and a maximum of 4, and when an individual comes in proximity of a different group it will join the new group with a probability of 0.9. Each group has a semi-circular shape with an average diameter of 2 meters. These parameters were chosen to simulate realistically the size, the inter-distance and the speed of movement of groups of individuals in the desired scenario, while maximizing dynamicity.

Simulator. Each individual wears a simulated device. We simulated the devices in accordance with the devices we used for our real world evaluation [45]. These nodes have an Atmel ATXMega128 CPU with 8k of RAM, 128k of Flash memory, and a Nordic nRF24L01+ wireless radio. To simulate the wireless network we used the OMNeT++[2] simulation engine with the MiXiM[3] framework for wireless and mobile network, the de-facto simulation environment for mobile ad-hoc and sensor networks [146]. The devices form an ad-hoc wireless network through an extension to the GMAC protocol [44]. In perfect conditions, the simulated transmission range is some 3.5 meters. For our purposes, we influenced these conditions in multiple ways through the simulator and protocol parameters.

Metrics. We extracted a series of proximity graphs directly from the coordinates of each mobility pattern and we consider these as our *ground truth*. We used a distance range of 3.5 meters to construct the proximity graph, as this is the ideal transmission range of our simulated devices.

[2] http://www.omnetpp.org
[3] http://mixim.sourceforge.net

To measure the performance of our proposed solution we compute the number of:

- *false positives* (FP): detections that are present in the measured proximity graph but not in the ground truth.
- *false negatives* (FN): detections that are present in the ground truth but not in the measured proximity graph.
- *true positives* (TP): detections that are present in the measured proximity graph and in the ground truth.
- *true negatives* (TN): detections that are missing in the measured proximity graph and in the ground truth.

We use these tests to compute two statistical measures of performance for binary classification tests. Sensitivity can be used to measure the ability of the test to identify positive results and is defined as $sensitivity = \frac{TP}{TP+FN}$. Specificity can be used to measure the ability of the test to identify negative results and is defined as $specificity = \frac{TN}{TN+FP}$. Intuitively, they measure the ability of the sensor to detect correctly proximity and the absence of it, respectively.

4.3.2 Setups

According to our approach, each sensor periodically broadcasts its ID through the radio, and a reception of such a transmission is considered as a detection. There are three main parameters we used to influence the behavior of the wireless network, and as such of the detections. We evaluated each of the 16 combinations of the three parameters over all the 10 mobility patterns, for a total of 160 simulation runs.

First, we can choose the number of *slots* used by the GMAC layer for the broadcasts. In GMAC, time is divided into a sequence of frames, and each frame is composed of a number of slots. The slot allocation algorithms of GMAC are based on Slotted Aloha [3], and are used to share access to the wireless medium. At a high level, the main difference between GMAC and Slotted Aloha is that GMAC exploits duty cycling. Slotted Aloha is an always-on protocol, whereas GMAC turns the radio on only for a small percentage of the total slots in order to save energy. Slots in which the radio is powered up are known as active slots, in contrast to inactive, or idle slots, where the radio is powered down. By controlling the number of active slots and their allocation strategy, we can control the number of "collisions" when multiple sensors are within the range d, and hence of loss of detections. In particular we experimented with the **4slots** setup, where four randomly allocated active slots are used (introducing a large number of collisions), and the

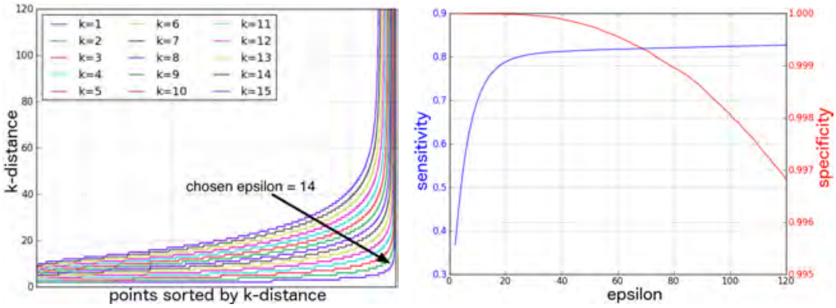
100slots setup, where each node has a slot assigned statically according to its ID (ideally avoiding collisions completely).

Second, *radio irregularity* can introduce variations in packet reception and it can be caused by factors related to the device and to the medium. Examples of the first ones include the antenna type, the transmission power, antenna gains, receiver sensitivity, receiver threshold and the Signal-to-Noise Ratio (SNR). Factors related to the medium are the medium type, the background noise, and some other environmental factors, such as the temperature and obstacles within the propagation media [169, 1]. While simulating some of the medium factors is more complicated, most of these factors are considered explicitly by the simulation engine [80], including those specific to the characteristics of our devices. Here, we concentrate in particular on the SNR, relaxing the assumption of a uniform disc of transmission and reception. We experimented with the **10SNR** setup, where a SNR threshold of 10 is used, and the **0SNR** setup, where the threshold is completely ignored.

Third, the *frame length* is expressed in time units, and it defines the periodicity with which nodes broadcast their ID. It defines the rate at which detections are collected. This rate acts as a sampling rate for our sensor, and it should be chosen according to the dynamicity of the behavior to be measured. For instance, it is desirable to use a frame length a number of times shorter than the minimum amount of time two sensors can be in each other's proximity. This allows to obtain potentially multiple detections during that interval and overcome possible loss of detections. We will show that increasing the sampling rate is very important to filter out noise. On the other hand, a higher sampling rate increases energy consumption, as the radio is used more frequently. There is a trade-off between energy efficiency and the sensitivity of the instrument. We experimented with the **1000ms**, **750ms**, **500ms**, **250ms** setups, which use a frame length as their name suggests.

4.3.3 Results

We ran the 160 runs of simulation and computed the sensitivity and specificity of the proximity graphs extracted from the raw measurements, and obtained through window-based smoothing and density-based clustering. Table 4.1 shows the results. For the raw measurement, the sensitivity simply measures the number of recorded detections, as it represents the naive approach that uses only set bits for the construction of the edges. For the raw measurement, two things should be noted. First, changing the number of slots and the allocation algorithm causes nearly 50% of



(a) k-distance plot (b) ε, sensitivity, and specificity

Fig. 4.3 (a) k-distance plot for scenario 4slots10SNR250ms. For readability, the plot is limited to points with a k-distance smaller than 120. The k-distance is expressed in frames, hence each frame here represents 250ms. (b) Relationship between ε, sensitivity and specificity for scenario 4slots10SNR250ms and minPTS = 1.

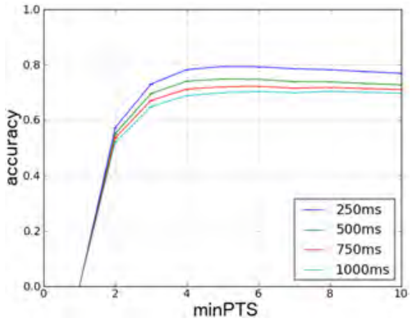
the detections to be missed due to collisions. Second, the change in SNR causes an additional miss of detections between 10% and 20%, depending on the number of slots used.

Looking at the results of our approach, one can notice that the two techniques do provide very similar results, as expected. They are able to reconstruct the detections missed due to collisions consistently, i.e., a very low standard deviation across the different runs for the same setup, and effectively, i.e., more than doubling the sensitivity in all the 4slots scenarios. On the other hand, our approach cannot improve the results for the 100slots scenarios. While this is understandable for the almost perfect conditions of the 100slots0SNR setups, one would imagine that our approach would reconstruct some of the data missed due to the SNR threshold. Moreover, the improvement introduced by our approach to both the 4slots scenarios brings the sensitivity close to the results for their corresponding 100slots scenarios, suggesting that most of the reconstructed data comes from the data missed due to collisions only. The reason is to be found in the very consistent and fast drop of coverage at the borders of the detection disc due to the SNR threshold in simulation. In a real deployment, the effect of the SNR threshold is to increase the probability of a radio transmission to be missed by another radio, as the sender reaches the border of the disc of the receiver. This probability increases quickly as the radio reaches the disc border, but still some transmissions should be received. What we experienced in simulation is that the radio passed from receiving everything to nothing. This basic on/off behavior is to be expected in absence of interference [109].

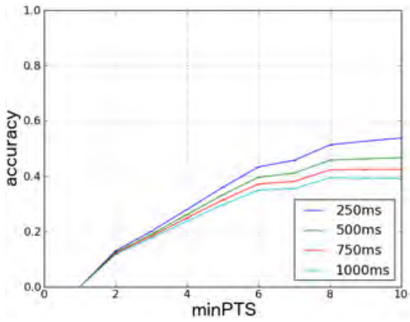
In effect, the SNR threshold causes the detection range to simply shrink. This hypothesis is confirmed by the absence of improvement when the frame length is decreased in the 10SNR scenarios. In fact, with a higher detection rate and without this on/off behavior, one would expect a higher chance that at least a few detections would make it through the disc border. Fortunately, this behavior is expected only in simulation, as it has been shown that the behavior of real links in low-power wireless networks, such as ours, deviates to a large extent from the ideal binary model used in several simulation studies. In particular, there is a large transitional region in wireless link quality that is characterized by significant levels of unreliability and asymmetry [164]. In the remainder of this chapter, we stick to the density-based clustering technique only, as it allows for an estimation of its parameters through the K-nearest neighbors analysis.

Choosing epsilon. To obtain the results presented in Table 4.1, we have carefully chosen the values for ε by using the k-distance plot as a reference. Figure 4.3(a) shows one of these plots and in particular the plot for our most realistic scenario 4slots10SNR250ms, for $k = 1, \ldots, 15$. For small values of k, the plot presents lines with a slope characterized by points with low k-distance values, and whose gradient changes once and very quickly towards the end. The points in the part of the curve with a low slope represent detections inside of clusters, and their k-distance value is the distance from their k-NNs. Note that all curves are consistently bimodal, meaning that the data presents a single density distribution, as expected and as required by DBSCAN. For readability, the plot presents only points with k-distance values under 120, but the data presents points with k-distance up to around 10000. These are the points that represent the distance between two detections that happened during two distinct and temporally distant moments. The k-distance is expressed in frames, so a k-distance of 1 represents 250ms.

To choose ε with $k = 1$, we concentrated on the elbow of the curve where the slope increases quickly, i.e., between k-distance 10 and 20 (for our experiments we chose a prudent value of 14 to preserve specificity). The points in this range represent detections that are in the sparse areas of the clusters or at their boundaries. Their k-distances are good candidates for our choice of ε. One has to be careful though, as a larger ε increases the probability to merge adjacent clusters. When this happens,



(a) Accuracy for 1% of noise



(b) Accuracy for 5% of noise

Fig. 4.4 Recognition of noise through different values of minPTS. Epsilon was chosen for each minPTS by reading the k-distance plot.

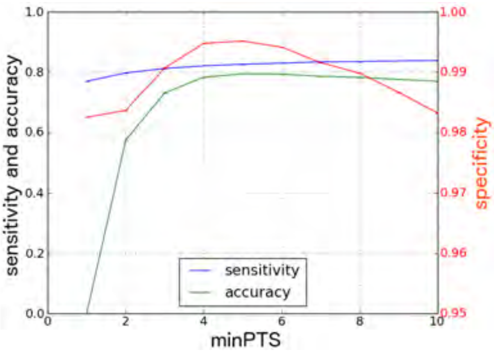


Fig. 4.5 Relationship between accuracy, sensitivity and specificity for 1% of noise.

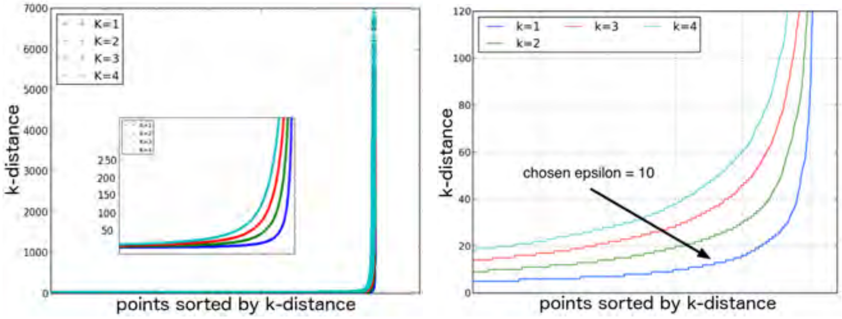
the algorithm sets also the bits that correspond to moments when the sensors were not in physical proximity. The consequence is an increase in the number of false positives. Although false positives do not affect the sensitivity of the measurement, they decrease the specificity. To investigate the relationship between ε, the sensitivity and the specificity we ran DBSCAN on the data that belongs to scenario 4slots10SNR250ms with different values of ε and $minPTS = 1$. The results are presented in Figure 4.3(b).

The results show that the sensitivity of the measurement increases with larger values of ε, as false negatives are turned into true positives by the algorithm. The sensitivity slows its slope after around $\varepsilon = 20$, as the optimal solution is reached. After this point, adjacent clusters start being merged, with true negatives being turned into false positives. This is evident by the specificity curve that starts decreasing around that same value of ε. These results show that good values of ε can be chosen by a correct read of the k-distance plot. Similar results can be obtained with larger values of $minPTS$. Note that the k-distance plot is often the only instrument available to estimate ε, as the measurement of sensitivity and specificity requires ground truth data.

Filtering out noise. Until now, we have kept $minPTS$ fixed to 1, so that also single detections could generate edges in the proximity graph. In a scenario where isolated detections can be generated by corruptions and interferences to the signal, a larger value might be necessary. This way, DBSCAN requires a larger minimum number of detections to establish a cluster, and this enables it to recognize noise. The downside is that short intervals of physical proximity risk to be filtered out as well, when the number of detections that characterize them is smaller than $minPTS$. This problem is particularly relevant for larger frame lengths, as they produce a lower detection rate and hence less detections for the same interval of time.

To investigate this hypothesis, we injected artificial noise into the data generated by the simulators. We ran three experiments where we set 1%, 5% and 10% of the bits in each bit string, respectively. For each pair of nodes, the bits were chosen uniformly at random. Each test was repeated 10 times and the results in Figure 4.4 and Figure 4.5 show the average values of accuracy in identifying the injected noise. The plots do present error bars, but due to the very low standard deviation they are not visible.

The results show that for 1% of noise the algorithm is able to recognize 80% of the noise already with a $minPTS$ of 4. For the experiment with 5%, 60% of the noise is recognized correctly with a $minPTS$ of 10. We do not present results for the experiment with 10% of noise as with $minPTS$ of 10 the accuracy was below 1%. These results validate our hypotheses. In fact, by specifying a larger value for $minPTS$ we are able to recognize an increasing amount of noise. Moreover, by increasing the detection rate the algorithm is able to recognize more noise, i.e., a difference of 10% between the 1000ms and 250ms frame length respectively. Note that, in particular with higher rates of noise, some of the randomly selected bits might be already set or belong to already existing bursts, hence not producing actual noise. This means that a 100% accuracy is an unrealistic result to reach. Overall, the results of these experiments show that, for realistic noise rates, the



(a) k-distance plot (with zoomed curve) of the ICT conference dataset

(b) (zoomed) k-distance plot for the controlled experiment

Fig. 4.6 Analyses of the bit strings extracted during two real-world experiments.

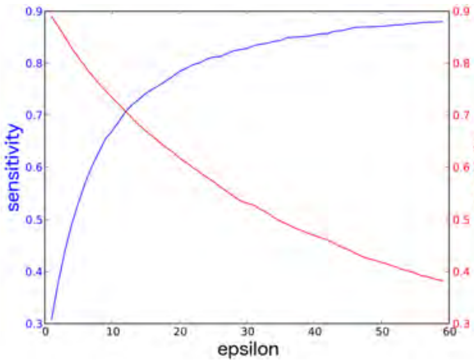


Fig. 4.7 Sensitivity and specificity study for the controlled experiment.

algorithm is able to recognize and filter out noise without a big impact on the sensitivity and specificity of the instrument. Also, they show the function of the detection rate with respect to the ability to filter out noise.

4.4 Evaluation in the real world

In this section we present the results obtained by applying the proposed approach to data collected during two real world social events.

4.4.1 Methodology

To explore the behavior of our approach on real data, we measured proximity information at two social events. In particular, we first conducted a controlled experiment in our laboratory, where we collected both proximity information and ground truth. Afterwards, we conducted a large-scale experiment during an IT conference attended by around 250 individuals. The devices were running the same software and protocol as the virtual devices in our simulations, and were hence using the GMAC protocol to broadcast their IDs every second to a short distance of 2-3 meters. Every device recorded the list of IDs received by the sensor at each second in the on-board storage unit, and we later collected this data for our offline analyses. The sensors were using 64 active slots and would choose randomly which slot to use to broadcast their ID. This combination represents a compromise between the 100slots and 4slots simulated setups. The devices were also broadcasting a second type of transmission. In fact, between two short-range transmissions, a long-range transmission was broadcast up to about 20 meters. This second broadcast contained the list of IDs received by each device during the previous frame, hence comprising only short-range detections. We captured these long-range transmissions through stationary devices called “sinks”. Sinks capture the radio transmissions exchanged by the devices, but do not transmit any data over the ad-hoc wireless network. We used this data to visualize at the events the evolution of the proximity graphs in real-time. Long-range transmissions were not used to detect proximity.

4.4.2 Evaluation against ground truth data

We conducted a first experiment in a controlled environment to collect ground truth data about physical proximity. For the experiment, 12 individuals participated to a cooperative social game in our laboratory. The experiment lasted about 45 minutes and was designed to replicate a conference hall scenario. The game required the individuals to form groups and interact to solve a quiz. At any time the individuals could switch groups, mimicking the way people switch conversations during a coffee break. During the experiment, the individuals were wearing our badge devices on their chest, in a similar way as a typical conference badge is worn. In addition to our badges, we used a real-time location system (RLTS) [4] to track the position of each individual at each time.

[4] Ubisense: http//www.ubisense.net

Moreover, we annotated the evolution of the formation of the groups through 5 video cameras and a human observer. From these two sources, we extracted two ground truth data sets.

Results. We performed the study of the sensitivity and the specificity of our proximity measurements against both ground truth data sets. The results were consistent with both ground truth sets but, due to space constraints, we report in Figure 4.7 only the results against the RLTS data, as it was a harder task. In fact, when our badge is worn on the chest, the body partially shields the transmissions, biasing the sensing area towards the front. On the other hand, the ground truth was extracted from the RLTS following the assumption of a uniform disc. As a result, some detections might be missed by our badges (e.g., when two individuals stand back to back), decreasing the sensitivity of the measurement. Reading the k-distance plot shown in Figure 4.6(b), we chose for ε a value of 10. The validity of the choice is confirmed by the study of the sensitivity and specificity for different values of ε shown in Figure 4.7. Precisely, using our approach with this parameter the sensitivity of the measurement improved from 0.309 (raw measurement) to 0.670, while the specificity decreased from 0.890 (raw measurement) to 0.733. Similar results were obtained against the annotated ground truth (sensitivity from 0.389 to 0.792 and specificity from 0.908 to 0.759). These results are strongly consistent with our simulations, as the sensitivity is more than *doubled* in the face of a decrease in specificity of only around 17%. This confirms the effectiveness of our approach also in a real-world scenario and make us confident of the quality of the measured proximity graphs.

4.4.3 Evaluation at an ICT conference

We conducted a second experiment at an ICT conference attended by around 250 individuals, 137 of which were wearing one of our devices. The conference presented six different tracks, and the track rooms were all scattered around the main hall of the event location. This main hall hosted the coffee breaks, the lunch banquet, and the poster session. We recorded proximity information for the whole event, but here we focus on the two hours between 12:00 and 14:00 when the lunch banquet and the poster session took place. No talks were given during this period, and the individuals were all gathered in the main hall. Due to the nature and the scale of the experiment, it was not possible to collect ground truth data.

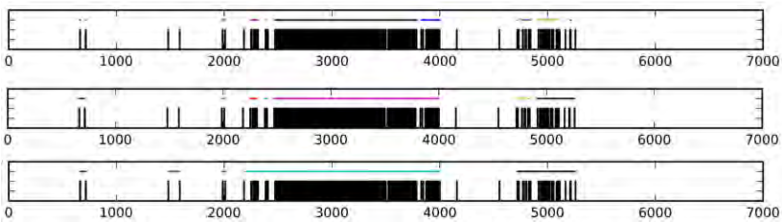


Fig. 4.8 The results of clustering detections using different values of epsilon and minPTS. Each plot represents the same proximity data between a pair of devices. (Top) epsilon = 30, minPTS = 2 (Center) epsilon = 60, minPTS = 4 (Bottom) epsilon = 120, minPTS = 2. Each vertical tick represents a set bit, the colored horizontal lines on top of the ticks represent the clusters those ticks have been assigned to. Ticks that are not beneath any line are classified as noise.

Results. After the experiment, we collected from each device the logs containing the list of IDs each device detected at each moment, and we converted the logs for each pair of devices to our binary format. Figure 4.6(a) shows the k-distance plot for the two hours we focus on. Also here, the k-distance plot presents a single density distribution, with a stable slope that increases only towards the right side of the plot. This result was expected and is consistent with the data previously obtained. We used this k-distance plot to choose the parameters to extract the proximity graphs from the dataset. Figure 4.8 shows the effect of the different values for the parameters used to cluster the detections between two specific individuals. Each plot represents the same bit string, and each vertical black tick in it represents a set bit in the raw data. The horizontal colored lines over the ticks represent the clusters the ticks have been assigned to. Each bit under these lines will result in an edge in a proximity graph. Ticks that are not beneath a colored line have been classified as noise. The top plot shows the results of clustering the ticks with $\varepsilon = 30$ and $minPTS = 2$, the center plot shows the results of clustering the ticks with $\varepsilon = 60$ and $minPTS = 4$, and the bottom plot shows the results of clustering the ticks with $\varepsilon = 120$ and $minPTS = 2$. These plots show how, by increasing ε, clusters grow by merging with adjacent clusters or by including isolated detections nearby. Also, they show how certain isolated detections can be classified as noise as a result of larger values of $minPTS$. Looking at these plots, it is possible to conclude that the two individuals were in physical proximity multiple times, with two larger intervals and a few shorter ones.

When designing an instrument that will be deployed to real environments, sources of mis-detections should be considered with high priority. During the experiment, we also measured the radio transmissions with a special device called an RF sniffer. This device scans a tunable band of radio frequency waves and attempts to interpret the observed signals as packets at high resolution. Where a normal device will observe nothing in the event of two messages colliding, an RF sniffer will observe some combination of the two messages generated by the overlapping of the two individual radio signals. As such, this device allows us to make some observations about the number of collisions/corrupt packets that would be experienced by devices in the vicinity of this sniffer. During the experiment our RF sniffer observed 277764 packets. Of these packets, 137077 were *valid* and 140687 (or just over 50%) were *corrupt*. These corrupt transmissions could be the result of many factors (e.g., interference from nearby WiFi devices), but the most likely cause is message collisions with other devices. These collisions will result in missed detections. In addition, it is likely that some corrupt packets (incorrectly) passed the CRC check, resulting in being accepted as valid data. The effect of these messages would be false positives, as the corrupt data would cause the device to detect other devices that are not truly at distance range, for example by modifying the transmitted ID with the ID of another device due to a flipped bit. These statistics show the necessity for a method such as ours to cope with these effects.

4.5 Discussion

We have shown how the application of data mining techniques to data collected through a number of wearable sensors can increase the reliability and the accuracy of physical proximity measurements. We have also shown how the setting of the parameters can influence sensitivity and specificity. In particular, trying to increase sensitivity exceedingly can cause specificity to decrease. Using an overly large value for ε can cause true negatives to be turned into false positives. Hence, there is a trade-off between the number of false negatives that can be turned into true positives and the number of true negatives that will be turned into false positives. It depends on the different applications and algorithms whether false negatives should be preferred to false positives and vice versa, putting the choice in the hands of practitioners. Regardless of this choice, one cannot assume a filtering of the data that yields to a perfect reconstruction of the measurement. In fact, algorithms and applications should be designed to cope with unreliable and inaccurate data, instead of assuming perfect measurements.

The presented approach assumes a single density distribution of detections that is valid for all pairs of devices at each time. While this assumption is reasonable for scenarios such as those described in this chapter, where the density distribution of individuals in space tends to be uniform and stable, more dynamic scenarios and longer measurements require a refinement of our approach. For example, individuals located in highly dense areas will experience a higher rate of false negatives and will require a larger ε. The same individuals, at a different time, might experience a lower rate in areas characterized by lower density. For this reason, for our future work we plan to extend our technique to consider data collected only during a limited amount of time, and adapt the estimate of ε accordingly. Moreover, we plan to investigate a stream-based technique that will allow us to consider only these most recent measurements in an *online* fashion. Ultimately, this more localized approach should enable the integration of the techniques directly on the devices.

4.6 Related work

Extracting proximity information about social behavior with on-body sensors is not a new idea. The Sociometer [32] is a wearable device that can measure social signals through IR sensors, microphones, Bluetooth, accelerometers, etc. As the focus of the Sociometer is to measure face-to-face interactions, proximity is sensed within a narrow frontal cone-shaped region of up to two meters. Given that people move around when they are talking, also the Sociometer suffers from bursty measurements. For this reason, the data extracted from the four IR sensors is processed with a Hidden Markov Model that is trained to learn patterns of IR signals over an annotated dataset. The focus on a frontal measurement and the dependence on annotated data pushed us to pursue a different approach. Hitachi's Microscope [159] is a device that provides similar features to those characterizing the Sociometer.

In [74] an approach based on RFIDs is presented. Similar to our approach, the device is worn on the chest of the individuals, and the devices periodically exchange their IDs through radio packets to detect physical proximity at close range. One of such detections accounts for an interval of proximity of 20 seconds. According to the authors, the devices operate with parameters that can assess physical proximity with a probability in excess of 99% over such time interval. As the parameters and the method are not described, it is difficult to identify the reliability of this assessment. Moreover, we aim at a more sensitive measurement than the one provided over those 20 seconds of potential false positives.

A different approach is followed for the iBadge [107]. The iBadge is a wearable device that can measure location, orientation and tilt, environmental settings, and speech. Orientation and tilt are measured from acceleration and magnetic sensors. The localization process is based on the location knowledge of fixed beacons (devices with known location) attached to the ceilings and the ability of the iBadge to accurately measure the distance between itself, the fixed beacons, and other iBadges in the room. The distance is measured by recording the arrival time difference of a simultaneous transmission of an RF signal and an ultrasonic burst. Using the distance measurements from multiple beacons in the room, iBadge estimates its precise 3-D location. Although iBadges cooperate in this localization process, the measurement of physical proximity is based on beacons. We consider this approach limiting and prefer a decentralized approach that does not need or produce location information.

4.7 Conclusions

In this work we have presented a method to sense physical proximity reliably. The method is based on a miniaturized device running an energy-efficient protocol for social ad-hoc wireless networks. The instrument can operate for weeks with one battery charge, making it suitable for long studies. The detections extracted from the network can be processed through a data mining technique that reconstructs missed detections and filters out noise. We evaluated both in simulation and in real-world experiments that the approach is able to increase the sensitivity of the sensor without lowering its specificity too much. The approach does not require a learning phase, but only a few parameters. We have provided a non-parametric method to choose the values for these parameters. The method provides a reliable measurement of proximity that is not bounded to the particular technology used to sense physical proximity, and can hence be used in a variety of scenarios and applications. Our approach is not specific to our devices or to a particular medium, but it is applicable to different technologies and protocols, as long as the technological layer underneath can map onto the proposed binary model.

5 Leveraging proximity sensing to mine the behavior of museum visitors

Without data, you are just another person with an opinion.

— **Andreas Schleicher**

5.1 Introduction

In the previous chapters we have discussed how person-to-person proximity detections can be leveraged to study crowd behavior. However, face-to-face proximity is not the realm of solely person-to-person relationships, but it can be used as a proxy to study person-to-object relationships as well. In fact, we face the objects with which we interact on a daily basis, like a television, the kitchen appliances, a book, including more complex objects like a stage where a concert is taking place. In this chapter, we focus on the relationship between the visitors of an art exhibition and its exhibits. We design, implement, and deploy a sensing infrastructure based on inexpensive mobile proximity sensors and a filtering pipeline that we use to measure face-to-face proximity between individuals and exhibits. Furthermore, we leverage person-to-object relationships to understand crowd behavior by identifying and predicting group behavior, which we visualize to gain insights into the behavior of the crowd.

Museum staff design exhibitions to educate, engage and entertain the visitors. Yet, museums rely on surveys and expensive observational

The contents of this chapter have been originally published in **"Leveraging proximity sensing to mine the behavior of museum visitors"** C. Martella, A. Miraglia, M. Cattani, M. van Steen - *IEEE International Conference on Pervasive Computing and Communications PERCOM* 2016 and **"Visualizing, Clustering, and Predicting the Behavior of Museum Visitors"** C. Martella, A. Miraglia, J. Frost, M. Cattani, M. van Steen - *Pervasive and Mobile Computing* 2016, and have been slightly modified to improve readability.

studies to collect coarse-grained information about the response of their visitors, with limits of scalability and bias. Central to the understanding of how visitors interact with an exhibition is the identification of which exhibits individuals stop at, for how long, and in which order. According to an ethnographic observational study of visiting styles at the Louvre museum, visitors can be classified into four classes according to their movements [140]. For example, one class of visitors follows a specific path spending a lot of time at almost all exhibits, while another class seems to have a specific preference for some known exhibits at which they spend a lot of time, ignoring the others. This classification has been used, based on activity logs coming from a digital museum guide, to help engage museum visitors and avoid information overload [82]. Characterizing and quantifying visitors behavior helps museum staff evaluate their curatorial decisions, reporting to stake holders and funders, and building data-driven marketing campaigns and applications.

The key to collecting this information is the fine-grained measurement of face-to-face proximity between visitors and exhibits. We call *positioning* the problem of identifying which exhibit an individual is facing at short distance. Existing approaches are not suitable for this problem. Indoor localization technologies estimate the absolute position of an individual in space, without measuring where the individual is looking at. This means that positioning an individual at the closest artwork within a given distance can produce false positives if the individual has her back towards the exhibit or if she is facing an artwork nearby. Moreover, only very few existing state-of-the-art techniques achieve, in optimal controlled environments, an error of less than 2 meters. With such errors, positionings of the visitor at a wrong exhibit, including those at the other side of a wall, would be frequent. Finally, these techniques often require expensive investments and rely on complex setup procedures each time exhibits are re-arranged.

Contributions. In this chapter, we propose a technique based on inexpensive and energy-efficient mobile proximity sensors, and a filtering pipeline to accurately position visitors at exhibits at all times. While we use radio-based proximity sensors, our filtering pipeline is not bound to any particular technology and does not require, for example, measuring distance from exhibits or expensive setups. In particular, we introduce a particle filter tailored to the problem of positioning, together with two smoothening filters that increase measurement accuracy. We use positionings to reconstruct the time spent at exhibits and the visitor path, defined as the ordered sequence of exhibits visited by the individual. We evaluate the approach with data collected from 182 volunteers during a real-world experiment. We show that by clustering this data we can identify and predict group behavior, such as common paths and patterns of time distribution at exhibits. Moreover, we show that by visualizing the behavior of the crowd we can help museum stuff gain a better understanding of the response of the crowd to their decision-making and planning.

The remainder of this chapter is organized as follows. First, we give an overview of our system, including sensing infrastructure, filtering pipeline, and data analysis. Then, we discuss related work regarding localization systems and museum technology. After describing our model and pipeline, we evaluate the accuracy of our measurements. We continue by presenting a number of visualizations of the data, and by applying state-of-the-art data mining techniques to the data. Finally, we conclude with a discussion about limitations and future work.

5.1.1 Motivation

Before deploying our infrastructure and running our data-collection experiment at the museum, we conducted a series of semi-structured interviews with the curator, the artistic director, and the head of business development to better understand their processes and needs. In addition to the staff of CoBrA, we also interviewed an independent exhibition designer who works regularly at CoBrA, and an administrator from the van Gogh museum who commissioned a large-scale observational study of that museum and has applied the findings to the re-organization of the permanent exhibition. Interview notes were coded and themes were identified. The analysis revealed three themes.

Inform design of future shows. Currently, museum staff designs exhibitions based on theories and experience about what pieces are popular and how visitors will move through the gallery spaces. For example, the artistic director at CoBrA believes that wall placement may impact visitors' movement through a show more than the displayed artworks, but expects that visitors also deviate from the ordering that curators envision. Both creative staff and those in charge of business development at CoBrA expressed a need for objective information to test such assertions and improve intuitive understanding about the gallery. The artistic director stated "If we learn more about that flow, [it] will have an impact on how we design, how we display the works of art."

Within a larger institution, the van Gogh Museum, parts of the galleries were redesigned based on the results of an observational study. The study had identified that the previous placement of certain famous artworks were causing congestion in parts of the gallery space. Moreover, less popular pieces were ignored when positioned close to more popular ones. Finally, the position of some divider walls confused some visitors about what path they were expected to follow in the museum, resulting in some parts of the exhibition being skipped completely. As we show in this chapter, similar observations — in fact corresponding visualizations — can be obtained with our sensor-based system, at a much lower cost.

Report to funders and potential partners. For the last ten years, the head of business development at the CoBrA has received requests from funders to demonstrate the value of shows. He stressed the need for "quantifiable information" about the experience of visitors. Museums track some metrics: ticket sales, museum shop sales and some periodic surveys, but these do not indicate how visitors interact with different artworks or level of satisfaction once inside of the exhibition. Specifically, the head of business development would like to know how much time visitors spend viewing particular pieces and how they move through the show. He wants to share these metrics externally in funding requests and inform the design of marketing campaigns. Finally, he found that, for example, identifying most popular artworks during the first weeks of a new exhibition could help the design of the marketing campaign, perhaps based on those pieces.

Increase participation. In the museum world, the staff of CoBrA reports a concern that visitors can remain passive and fall into the "museum shuffle", that is, visitors walk from piece to piece mindlessly, spending a few seconds at each piece before moving on. Museum staff wants to find ways to know when this occurs such that they can intervene and change behavior. The ultimate goal is to avoid this passive mode and enhance engagement with the content to increase the exhibitions "impact" beyond the visit itself. The interviewees' main objective for a museum visit is for visitors to learn something. To do so, they want to "get in touch" with visitors using different methods of understanding the visitor experience. The artistic director acknowledged current methods were not providing the right type of information: "we know them in statistics [...] but we do not know about the quality of their experience".

The results of these interviews were used for the design of our data collection and data analyses.

5.2 Overview

We conducted a 5-days experiment spread across 2 weekends at the CoBrA Museum of Modern Art (CoBrA). Our data collection focused on the temporary exhibition entitled “The Hidden Picture”, a curated sample of the corporate collection of ING. The exhibition was displayed in the dedicated open space at the top floor of the museum. The space is configurable, and divider walls were used to separate the space into 6 “open rooms” dedicated to different themes. The overall space was about 100 meters long and 25 meters wide, with a ceiling reaching about 5 meters, while divider walls were some 3.5 meters high.

Rooms `1` and `2` focused on figurative art, rooms `3` and `4` mostly on abstract art, room `6` on pieces inspired by nature, for a total of 60 pieces. The pieces varied in size, style and medium, including photos, paintings, sculptures, videos, and an installation with a cage hosting a living chameleon. None of the pieces were highly famous, and were hence appealing the visitors based on immediate reaction rather than on prior knowledge. Of the 60 pieces, we instrumented 45 exhibits with our sensing infrastructure.

5.2.1 Data-collection architecture

We designed a system based on inexpensive radio-based proximity sensors. Our sensing solution is compliant to the Zigbee standard and it can be implemented for example through Bluetooth low energy (BLE) beaconing, available in modern smart phones. To give us freedom to investigate our solution, instead we deployed ad-hoc devices running a duty-cycled MAC protocol [142] that allows us to run our system for weeks with a single battery charge.

The sensing infrastructure comprises *mobile* devices and *anchor* points (or simply anchors). Mobile devices are sensor nodes worn by the visitors. They are attached to a lanyard worn around visitors’ neck and hang on the chest. Due to the shielding effect of the visitor’s body, the radio communication range is steered to the front with a controlled angle of around 90 degrees and some 2-3 meters of distance. Anchors are sensor nodes positioned at the base of each exhibit. We installed anchors inside of enclosure aluminum boxes designed to shape the communication range to approximately 60 degrees and 2-3 meters of distance. With this setup, mobile devices and anchors can communicate only when the visitor is facing an exhibit.

Every second, anchors transmit through the radio a unique anchor identifier (*AID*) that is received and timestamped by mobile devices within range. We consider the reception of an AID by a mobile device a *proximity detection*. Note that our sensors do not measure radio signal strength (i.e., RSSI). While it does not enable us to measure distance between points, it allows a cheaper and more energy-efficient solution. Every second, mobile devices transmit the list of detections received during the previous second together with their unique mobile-point identifier (*MID*) to a longer range of approximately 100 meters, which are received by one or more *sinks*.

Sinks are computers that receive mobile devices transmissions through the same type of sensor node used for anchors and mobile devices, and store the timestamped lists of detections in a central repository. Sinks are installed in various areas of the exhibition space to ensure full coverage and some degree of overlap. Note that due to the overlap of the areas covered by the sinks, mobile devices transmit their messages together with a randomly generated number that we use together with timestamps to remove duplicate detections from the database.

When a mobile device is handed out anonymously to a visitor, the visitor is assigned a unique user identifier (*UID*) that is associated to the corresponding MID. Each visitor check-in and check-out times are stored together with the UID-MID mapping. Our raw data database comprises this mapping and the list of timestamped detections collected by sinks.

5.2.2 Data-filtering pipeline

The raw database of proximity detections is characterized by a number of shortcomings that obstacle a direct use for visitor positioning without prior filtering. Accurately computing which artwork, if any, a visitor is facing at each second based solely on raw data is not possible for a number of reasons, related to the irregularity of wireless communications and the imperfect steering of our enclosures.

We designed a data filtering pipeline to estimate, for each second of a visit, at which artwork a visitor is positioned and the sequence of exhibits that defines such visit. The pipeline comprises three steps. First, we filter the data with a *particle filter*. We have developed a technique based on particle filters that takes into account the topology of the exhibition room, the placement and directionality of the anchors, and the movement of the visitor. Note that our technique positions the visitor at exhibits and does not compute absolute coordinates, like traditional techniques based on particle filters designed for localization.

Second, visitor-exhibit mappings are further filtered with a *density-based filtering* algorithm that corrects occasional artifacts introduced by the particle filter. While the particle filter drastically increase positioning accuracy, there are still occasions where a large number of missing detections can cause gaps in positioning data, making it appear as a visitor would return at an exhibit repeatedly in a short window of time (with some seconds in between where the visitor appears having left the exhibit either for the center of the room or, more rarely, for an exhibit nearby). The filtering algorithm fills these gaps and ignores spurious positioning data.

Third, we extract through a *majority voting* filter the sequence of AIDs at which the visitor was positioned. The filter scans a visitor positioning data to detect transitions from exhibit to exhibit, and disambiguating situations where a visitor may appear facing two exhibits at the same time. During this step we also reconstruct the *path* followed by the visitor, that is the sequence of stops at exhibits followed by the visitor.

5.2.3 Data-analytics applications

Once raw data has been processed by our pipeline, we obtain for each visitor a vector r_v of N elements, each representing the number of seconds spent at each of the N anchors/exhibits, and a sequence s_v of AIDs to represent the path followed to visit the exhibits. If an exhibit was never visited by the volunteer, the corresponding element of r_v will contain a value of 0, and the corresponding AID will be missing in s_v.

We can leverage these two data structures to mine the behavior of all the visitors and discover behavioral patterns, like popular rooms and exhibits, common paths followed through the exhibition. Furthermore, if group behavior did emerge during the exhibition, one can leverage historical data of past visitors to predict the behavior of future visitors, from time spent at exhibits to their satisfaction.

5.3 Related work

Our work is closely related to the topic of indoor localization. A recent evaluation of 22 indoor localization mechanisms [94] provides us several insights on how state-of-the-art localization techniques could perform in our scenario.

Among all localization mechanisms, only three [118, 14, 90] achieved an error of less than 2m, while only half achieved an error of less than 3m. Moreover, localization errors increase significantly (both in terms of average and deviation) at the edges of rooms and in hallways, were most museums have exhibits. For our application, this is quite a significant error since the distance between exhibits is usually just few meters. Differently from localization techniques, our mechanism provides a deployment density that matches the placement of exhibits. In other words, by design we tailor our positionings where needed.

Moreover, because museums often have exhibits on both sides of walls, this lack of accuracy produces an even larger error when used for positioning (e.g. positioning a visitor at a painting in the next room or even worse on the next floor). Our combination of enclosures and a tailored particle filter allows us to focus on face-to-face proximity and minimizes these errors. Furthermore, localization systems are usually evaluated in controlled settings and do not account for the variability introduced by people or by changes in the furniture setup, both typical aspects of temporary exhibitions. In more realistic conditions, the localization error of the tested indoor techniques increase by approximately 1.5m to 4m.

There are mainly two types of system. Infrastructure-free mechanisms [118, 90, 170] exploit the existing Wi-Fi access points and require a lengthy calibration phase (*fingerprinting*) that must be repeated every time the environment significantly changes. In case of a museum, this means every time the exhibition changes. Infrastructure-based approaches, on the other hand, usually require a one-time deployment of a hardware infrastructure that can be expensive and that could affect other practical issues such as aesthetics and safety certifications.

Instead of fingerprinting the radio environment, which is susceptible to changes in the environment, most infrastructure-based techniques exploit the propagation speed (*time-of-flight*) of radio [4, 9] and sound [87, 76] to estimate the distance of visitors from the anchor and triangulate their position. Interesting to note, the only localization technique able to achieve sub-meter accuracy [118] is not based on time-of-flight, but on the signal's phase offset, and took several years (5) of development to provide such accuracy. Being based on a proprietary hardware, this technique is difficult to generalize and apply, for example, to smart phone platforms.

Our technique, on the other hand, is generic and based only on face-to-face proximity, that is fuzzy and hard to predict by nature. In particular, since our mechanism does not require to sense the signal characteristic such as the strength or the phase offset, it can be applied to several protocols (e.g. ZigBee, WiFi, Bluetooth, etc.) and even different mediums (e.g. radio, sound and light).

Few sensor-based systems have been used to study the behavior of visitors of art exhibitions, in relation in particular to their movements. Early attempts made use of indoor localization systems based on Bluetooth data collected from mobile phones, to trace the movement of visitors between rooms [162]. This data can be used, for example, to support multimedia guides [149, 22], but it captures only which room (or part of it) an individual is visiting. More recently, data coming from indoor localization and physiological sensors has been used, together with entrance and exit surveys, to study the cognitive reaction and social behavior of a number of individuals in an exhibition [78].

A similar device to measure position and spatial orientation (i.e., through a compass) of the individuals has been used to study the behavior of visitor pairs in a museum. The study presents a system to classify pairs early in the visit into one of six classes, to provide socially-aware services to the pairs, for example to increase their engagement with the exhibition [42]. This study focuses on the interaction between the individuals and does not attempt to position the pairs at exhibits.

To summarize, none of these approaches tackles the problem of fine-grained face-to-face positioning, and rely either on coarse-grained room-level positioning or on absolute localization without attempting to position visitors at exhibits.

5.4 Model

We consider a visitor wearing a mobile device v for a duration of T seconds. The exhibition comprises N exhibits, each instrumented with an anchor $a_i \in A = \{a_1, a_2, \dots, a_N\}$. For each visitor v we represent the set of proximity detections as a $N \times T$ matrix D_v, where $D_v(i, j) = 1$ if and only if the mobile device v detected anchor a_i at time j. $D_v(*, t)$ refers to all detections collected at any time t, and $D_v(i, *)$ to all detections of a_i. Similarly, we define a positioning matrix M_v as a $N \times T$ matrix, where $M_v(i, t) = 1$ if and only if visitor v was facing exhibit a_i at time t within a distance smaller than d (i.e., the sensor detection range, in our case 3m.). The definition of M_v is analogous to the definition of a series of proximity graphs [97] (see Chapter 3), with the additional constraint of a visitor being in proximity of maximum one anchor at any time t.

Note that there can be times t such that v is not positioned at any exhibit. These are the times when a visitor is walking around the museum or too far to be detected.

5.4.1 Problem definition

Our goal is to compute the positioning matrix M_v from the detection matrix D_v. Ideally, if v was facing a_l from second i to second j within distance d, we would have $D_v(l,k) = 1$ for $i \leq k \leq j$ for *only* a_l. In other words, D_v would contain detections for the whole duration of the face-to-face proximity between v and a_l, and there would be *only one* a_l for a given time k such that $D_v(l,k) = 1$. With these perfect conditions, matrix D_v would contain a continuous stream of detections between v and the anchors, and it could be directly used as positioning matrix M_v. However, in practice this is not possible for the following reasons.

First, the database is missing detections. Transmissions between anchors, mobile devices, and sinks can be lost due to many factors, such as message collisions and low signal-to-noise ratios [169, 1]. In other words, while a visitor faces an exhibit, sinks collect a *bursty* stream of detections with *gaps*.

Second, although we control the communication range of the anchors with our enclosures, it is still possible for a mobile device to detect multiple anchors at the same time. A typical scenario is when two anchors are positioned close-by and the visitor stands in an area of overlap between the two anchors' transmission ranges. Looking at the instantaneous raw data does not allow disambiguation of these occurrences. Note that this problem could occur also between anchors positioned at the two sides of a wall, causing a visitor to be in proximity of exhibits installed in two different rooms at the same time.

Third, a mobile device can detect an anchor at a distance larger than the expected transmission range, either due to a corrupted AID in the anchor message (typically caused by a collided message passing a CRC), or due to some tunneling effect in the wireless transmission. This is perhaps the problem occurring less frequently.

We hence define our pipeline as the set of operations to filter and transform D_v into a M_v for any given visitor v.

5.4.2 Particle filter

We designed and implemented a filter based on particle filters to tackle two challenges related to our positioning problem: (i) due to unreliable wireless communication, some detections could be missing, and (ii) due to the multipath effect, a mobile device could detect multiple anchor points, even far away and on the other side of a wall.

Particle filters have been successfully used in localization to estimate the absolute position of individuals with unreliable sensors [56, 63]. For localization, usually a mobile sensor communicates with a few anchors installed at known locations. It is assumed that the sensor can communicate with all, or a majority of, anchors from all positions and directions, and that the sensor can measure distance from these anchors, for example through signal strength or time-of-flight.

Our setup does not match these assumptions, as we deploy many anchors that communicate only directionally and at short range, without the capability to measure distance. Moreover, we would expect, and to a certain extent desire, mobile devices to detect only one anchor at each time, if any, that is the one in front of the visitor. In other words, our system was designed for the problem of positioning, therefore we need to design a particle filter that reflects the characteristics of our setup and problem.

The filter requires topology information about the exhibition. In particular, a set of anchors $A = \{a_1, a_2, \dots, a_N\}$ each characterized by a position x, y and an orientation α, and a set of walls $W = \{w_1, w_2, \dots, w_M\}$, each defined as a segment between two points, to describe the layout of the exhibition space. We define the set of particles $P = \{p_1, p_2, \dots, p_K\}$, each defined by a position x, y and a weight w (the likelihood of a particle to represent the actual visitor's position). Initially, particles are distributed uniformly at random across the layout of the exhibition space.

Given a detections matrix D_v, the particle filter comprises four steps that are executed for each time t of the visit, with $0 \leq t < T$.

- **Estimation**: We compute the likelihood of each particle's estimate (i.e., its position) given the measurement at time t, that is the set of detections in D_v contained in the t-th column . For each particle p, its weight is computed using the likelihood function $w = \Phi(p, D_v(*, t))$. More details about this function are given later in the text.
- **Positioning**: We estimate the position of v by computing the weighted average among the particle coordinates (i.e., the centroid) and find the closest anchor a_i. We then set $M_v(i, t) = 1$, unless the variance of the particle coordinates from the centroid is larger than a threshold, meaning that the confidence of the estimate is low.

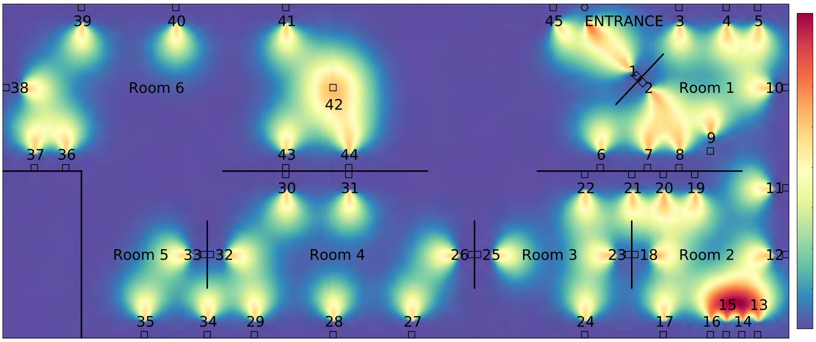


Fig. 5.1 The landscape of the multivariate gaussian kernel used to compute the likelihood function Φ. For displaying purposes, we assume all anchor points were detected at time t. Note that the likelihood function takes into account angle and distance from anchors, as well as overlap between detection ranges.

- **Re-sampling**: We create a new set of particles by drawing with replacement from the current weighted set of particles. While drawing particles from the set, we favor particles proportionally to their weight (i.e., their likelihood). As a result, particles with higher likelihood are picked more often than particles with lower likelihood. If all particles are improbable, we generate new ones at random.
- **Movement**: We move particles at walking speed in random directions, avoiding illegal moves, such as walking through walls.

An important component of our system is the function to compute the likelihood of a particle's estimate. Intuitively, because our sensors are steered to measure face-to-face proximity, we want a particle to have higher likelihood if it is positioned in front of a detected anchor and if it is at close distance. We define $\Phi(p, D_v(*,t))$ as a multivariate function based on a Gaussian kernel with maximum likelihood at 1 meter distance and at 0 degrees angle from the anchor (exactly in front). The likelihood is defined as 0 for distances beyond 3 meters and angles larger than 30 degrees. When multiple detections are present, we mix the likelihood functions of the involved anchors by summing their values. In those cases, the likelihood function computes a multi-lateration between the detected anchors.

For displaying purposes, Figure 5.1 shows the landscape of Φ assuming that all anchors are detected at the same point in time. In case only one anchor is detected, one can imagine the landscape to have values larger than 0 only in front of the anchor. Note that artwork `42` was a cage, positioned far from the walls, of which visitors could walk around. For this reason, we placed the sensor inside an enclosure that allowed detections from any angle, still within the same distance as defined for the other artworks. The function models directionality and range of both mobile devices and anchors, leveraging the radio directionality due to the enclosures.

The filter keeps positioning a visitor at the last exhibit during gaps of missing detections. As soon as particles spread too much due to their random movements, the filter assumes the visitor is far from any exhibit. Similarly, spurious detections from anchors far away or at the other side of a wall are ignored as it takes time for particles to reach them (e.g. without crossing walls), and hence for their likelihood to be affected by those anchors.

5.4.3 Density-based filter

The particle filter computes the positionings of a visitor regardless of missing, wrong, and ambiguous detections. The resulting matrix M_v contains for each anchor a series of bits that tells us whenever a visitor was facing an anchor at a certain time. While smoothened by the particle filter, the series in M_v can still present gaps, for example, during short periods where particle confidence was too low (wrongly, due to many missed detections), or when a visitor was wrongly associated to a nearby exhibit. Even though they are rare, we want to remove these artifacts by further smoothening M_v through a density-based filter. In principle, the density-based filter acts analogously to a low-pass filter implemented through a sliding window, but it is able to compute the optimal values for the parameters corresponding to window size and threshold.

The density-based filter [98] (see Chapter 4) first analyses all the series in M_v to compute the k-nearest neighbors statistics for each positioning, and it then uses these statistics to automatically identify bursts of bits through the density-based clustering algorithm DBSCAN [55]. Each identified cluster is effectively a period of time when the visitor was continuously facing an exhibit. Once clusters are identified, we can fill the gaps within the clusters. Note that, at the same time, the clustering algorithm classifies positionings that are isolated and not part of any cluster as noise. The result is a new positioning matrix M'_v that is effectively a smoothened version of M_v.

Note that, differently from M_v, in M'_v visitor v can be positioned at multiple anchor points for the same time t, as series are filtered independently. We solve this problem in the next, and last, step of the pipeline.

5.4.4 Majority-voting filter

We define a majority-voting filter to disambiguate those times in M'_v when we position v at more than one anchor (i.e., those columns of M'_v that have more than one row with a value of 1). The majority-voting filter looks at a window of duration of L seconds ahead of, and including, t to decide at which anchor to position v. The filter decides by choosing the anchor with the largest number of positionings in that window of time. Formally, we let N windows $w_{j,L,a_i} = \langle M'(i,j), M'(i,j+1), \ldots, M'(i,j+L)\rangle$ slide in parallel over the series (i.e., rows) in M'_v, with $0 \leq j < T-L$. We break ties by picking the anchor chosen at the previous slide, or at random when no anchor was chosen at the previous slide (i.e., when all windows contain all zeros). While we filter M'_v we create a new positioning matrix M''_v and a path sequence s_v of anchor AIDs that represents the order used to visit the exhibits. We consider v to have transitioned to a new exhibit when the filter positions v at an anchor different from the previous.

To summarize, for each visitor v, the pipeline outputs (i) the positioning matrix M''_v representing at which anchor v was positioned, if any, at each time t, and (ii) the sequence s_v of anchors AIDs that represents in which order v visited the exhibits. Note that not necessarily all anchor are present in s_v and also that an anchor AID can appear multiple times, though not consequently, in s_v. Finally, we can compute a vector r_v of length N, that represents the number of seconds spent by the visitor facing each exhibit. More precisely, we compute the values of r_v as $r_v[i] = \sum_{t=0}^{T} M''_v(i,t)$, for $0 < i < N$. We will use r_v and s_v in Section 5.7 to identify group behavior.

5.5 Evaluation

5.5.1 Methodology

We organized two experiments at CoBrA to evaluate our model and infrastructure. In a *controlled* experiment, we scripted 28 visits and asked volunteers to follow the instructions through the script with a timer. The script defined a visit as a sequence of stops at exhibits, each characterized by a time of arrival and a time of departure from each exhibit. Volunteers were asked to stay at some 1-2 meters of distance and facing the exhibit for the whole duration of each stop. For this experiment, we focused on Room 4 only. While we focused on a subset of the exhibits, we kept all anchors on at all times, including those at the other side of the walls.

In a *real-world* experiment, we asked the visitors of the museum to volunteer in the experiment by wearing one of our sensors during their visit. Volunteers were not instructed or scripted in any way, and could move freely in the exhibition space for the whole duration of their visit. A total of 182 volunteers decided to participate, spread over the 5 days of duration of the real-world experiment. Two human observers collected *ground truth* positionings for 19 volunteers, by annotating arrival and departure times at each exhibit (corresponding to M''_v), and the order used to visit exhibits (corresponding to s_v).

Setups. A the end of the experiments, we processed the data collected by the sinks in the central repository. We utilized the same sensing infrastructure, i.e., the sensors and the enclosures, as well as the same filtering pipeline for both experiments. For the particle filter, we utilized 1000 particles and set the particle speed to 1m/s. Density-based filter parameters were chosen as described in [98] (see Chapter 4) by choosing the knee point of the k-distance plot, corresponding to $\varepsilon = 15$ for $minPts = 2$. We chose a window size $L = 10$ that we chose empirically as it would maximize accuracy of both positionings and paths, though results did not vary significantly in the interval $[5, 30]$.

Metrics. To measure the performance of our solution at the task of positioning visitors at the exhibits we compute the number of:

- *False positives* (FP): positionings that are present in the measurement but not in the ground truth.
- *False negatives* (FN): positionings that are present in the ground truth but not in the measurement.
- *True positives* (TP): positionings that are present in the measurement and in the ground truth.
- *True negatives* (TN): positionings that are missing in the measurement and in the ground truth.

We compute these values by comparing M''_v and the ground truth for each annotated (or scripted) volunteer. We use these tests to compute two statistical measures of performance for binary classification tests. Sensitivity can be used to measure the ability of the test to identify positive results and is defined as $sensitivity = \frac{TP}{TP+FN}$. Specificity can be used to measure the ability of the test to identify negative results and is defined as $specificity = \frac{TN}{TN+FP}$. Intuitively, they measure the ability to correctly estimate positioning and its absence, respectively. Balanced accuracy, used in cases of unbalanced classes such as ours (where true negatives are much more frequent), is defined as the arithmetic mean of sensitivity and specificity.

To measure the performance of our solution at the task of computing paths, we compared paths extracted through our method with paths extracted from the ground truth. To this end, we use two types of metrics: *sequence-based* and *coordinate-based* metrics.

Sequence-based. We used two metrics designed to compute *similarity* between sequences. The first metric is the Jaro-Winkler (Jaro) [151] similarity metric, which is used to compute string similarity. Jaro is a type of string-edit distance that considers explicitly the mismatch in the order in which elements appear in two sequences (an operation called transposition) and how far mismatching elements are in the sequences. Intuitively, inverting in s_v the order of two exhibits that are nearby in the sequence is less penalized than inverting two exhibits far from each other in the sequence (e.g. $jaro(``ABCDE'', ``ABCED'') > jaro(``ABCDE'', ``AECDB''))$. The second metric is the Ratcliff-Obershelp (Sequencematcher) [112] matching metric, which is used for pattern recognition and is less forgiving when inverting nearby elements. Sequencematcher matches recursively the elements in the longest-common-subsequences between two given sequences. Both metrics compute a value between 0 and 1, with 1 representing perfect similarity and 0 representing no match.

Coordinate-based. We used a metric designed to compute the *distance* between two sets of coordinates. Hausdorff distance measures how far two subsets of a metrics space are from each other. Intuitively, two sets of coordinates are close if every point in one set is close to some point in the other set. More precisely, Hausdorff is defined as the longest distance from a point in one set to the closest point in the other set. While Hausdorff distance is originally defined on the longest distance, often the mean and median distance are also used to gain a better picture of the distance between the two sets. We generate a set of coordinates from a sequence s_v as follows. We create a mesh by splitting the layout of the exhibition space in $1m^2$ cells. Mesh edges represent cell edges, and mesh vertices represent cell vertices. In addition to cell edges, we also connect vertices through edges representing cell diagonals. We do not allow edges crossing walls. Every vertex in the mesh has an

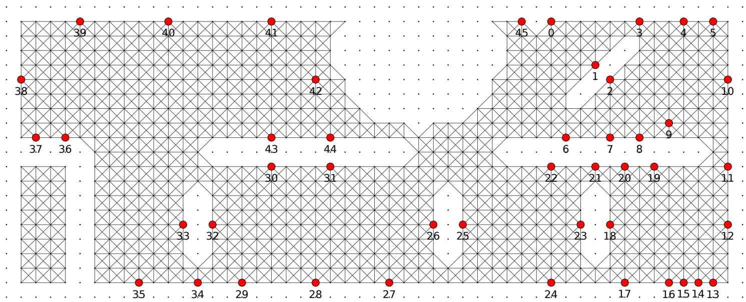


Fig. 5.2 The mesh used as layout to compute paths between paintings. The planimetry is split in cells, each one mapping 1 square meter of the layout space. Red dots labeled with a number represent artworks. Walls are represented as disconnected dots in the mesh. The circular gap in the upper-central part of the mesh is due to a non-walkable atrium.

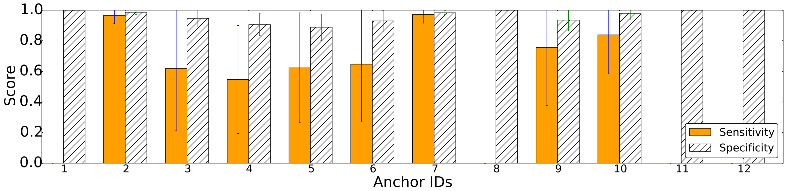


Fig. 5.3 Anchors sensitivity and specificity values for the controlled experiment across annotated/scripted visitors. Some anchors have no sensitivity because no stops were scripted at those anchors (they were positioned at the other side of walls of anchors involved in the experiment).

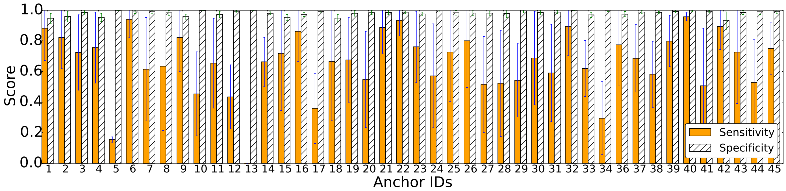


Fig. 5.4 Anchors sensitivity and specificity values for the real-world experiment across annotated/scripted visitors.

associated coordinate that depends on its position when overlayed on the exhibition space layout (with the $(0,0)$ vertex being positioned at the top-left of the layout). Figure 5.2 shows the mesh. We compute the shortest paths between each pair of adjacent elements in s_v and we concatenate the list of coordinates associated with the vertices in the shortest paths. This way, we obtain a representation of the path in space

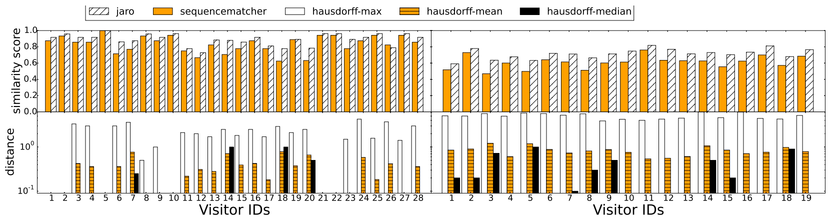


Fig. 5.5 Sequence-based similarity values (top) and coordinate-based distance (bottom) for the controlled experiment (left) and the real-world experiment (right). We plot coordinate-based distance in log-scale.

that visitor v would have followed while performing s_v, had she followed the shortest one.

While we do not expect the visitor to have walked precisely through those coordinates, we used also a coordinate-based metric to better measure the impact of errors in s_v. A missing or wrongly added anchor in s_v may result in a single string edit operation of Jaro that does not capture whether the visitor was positioned far away in space from the correct anchor, for example, on the other side of a wall compared to the anchor 1 meter away. Introducing a coordinate-based metric allows us to understand the impact of our errors in relation to the placement of the anchors in the exhibition space. Moreover, this metric allows us to understand the accuracy of path visualizations, as two subsequences with low Hausdorff distance (e.g., if we miss an anchor by positioning the visitor at the anchor 1 meter away) look very similar, often enough for a visualization, regardless of having potentially a low sequence-based similarity.

5.5.2 Results

Positioning. Figure 5.3 and Figure 5.4 present the mean values and standard deviations of sensitivity and specificity for each anchor. The average sensitivity across the anchors for the controlled and real-world experiments was 0.73 (std 0.02) and 0.61 (std 0.04) respectively, while the average specificity was respectively 0.944 (std 0.002) and 0.981 (std 0.001), for a balanced accuracy of respectively 0.84 and 0.79. Note that the pipeline increases balanced accuracy by 42% in the case of the real-world experiment, and by 63% in the case of the controlled experiment, compared to using raw detections as positionings.

We can notice that most errors are caused by false negatives, as the values of specificity are very close to the maximum value of 1.0, while sensitivity values are smaller and vary more. This can be expected to be caused by missing detections, which are more likely than false positives. Moreover, reaching very high sensitivity may be extremely difficult due to the method used to collect ground truth. As visitors face exhibits for short periods of time (on average around 15 seconds), a small annotation error can impact substantially the sensitivity metric. Observers were asked to start timing a positioning at an anchor when the visitor was facing the exhibit from a distance of some 2-3 meters, but it is difficult for the observer to identify the exact moment the visitor is at range and facing the exhibit. An error of 1-2 seconds about arrival and departure times by the observers can mean missing $10-20\%$ of the true positives.

Moreover, often visitors moved during their time at exhibits, getting close and further from the exhibit, temporarily facing somewhere else to discuss with an another visitor or to approach an information sheet attached to a wall nearby the exhibit. It is almost impossible for the observers to include all these fine-grained elements in their annotations. This hypothesis is confirmed by the increased accuracy obtained with the controlled experiment, though also in the case of scripted visits it is difficult for visitors to consistently approach and depart the detection range correctly with second-level precision as defined by the script.

Some sensors in denser areas reach often either very low or very high sensitivity, while variance is lower in areas of lower density. This can be an expected effect of the last stages of the pipeline, where smoothening filters favor certain anchors in a winner-takes-all fashion, perhaps due to a favoring conditions of the enclosures position and orientation (e.g. exhibits `5`, `13` and `34`). This happens less frequently between anchors positioned more distantly from each other.

To investigate the impact of the variability in the sensitivity we ranked the artworks by the amount of time spent in total by all annotated visitors facing them. We ranked exhibits based on the measurement and on the annotations. Intuitively, we would expect the two rankings to be similar if the measurement was accurate enough regardless of variance in missed positionings (i.e., we miss some positionings but we can still order the exhibits by time correctly). We computed Spearman rank correlation between the two rankings to see whether we can capture the relative relationships between anchors. The correlation values for the whole ranking was 0.413 with a p-vaule of 0.01. We then computed correlations for the top N exhibits for $N = 5, 10, 15, 20$, yielding respectively correlations values 0.996 ($p = 0.0002$), 0.768 ($p = 0.009$), 0.65 ($p = 0.008$), 0.544 ($p = 0.005$). These results show that errors are accumulated at the tail of the ranking, where time spent at exhibits are more similar and hence easier to mistake. We would expect to converge to better results with a sample larger than the 19 annotated individuals.

Paths. Figure 5.5 shows sequence-based similarity and coordinate-based distance values for both experiments, with values for each annotated or scripted visitor. The average Jaro similarity across visitors for the controlled and real-world experiments was 0.879 (std 0.08) and 0.716 (std 0.05) respectively, while Sequencematcher similarity was 0.814 (std 0.09) and 0.613 (std 0.05) respectively. The average sequence lengths were respectively 8.147 for the controlled experiment and 36.316 for real-world experiment. The measures show that we are able to accurately reconstruct path sequences, even when they are long. Moreover, higher values of Jaro similarity confirm that sometimes we invert nearby anchor points in sequences.

Regarding coordinate-based distance, the average (max, mean, median) values of Hausdorff distance for the controlled experiment and the real-world experiment were (1.186, 0.317, 0.137) and (5.002, 0.825, 0.245) respectively. Mean and median sub-meter error show that we consistently position the visitor correctly (or less frequently at an anchor nearby), while a 5 meters maximum distance suggests that worst errors cause a positioning of a visitor at a "walking" distance of 5 meters.

5.6 Visualizing visitor behavior

In this section we present a number of visualizations that we have developed to show insights of visitor (group) behavior present in the proximity data. While our work is complementary to the existing related work regarding attraction power [85], it introduces a richer understanding and presentation of paths, moving from existing work that pictures only exhibit-to-exhibit transitions to full sequences of exhibits and their frequency (i.e., paths).

5.6.1 Artworks and rooms popularity

Based on the findings of our assessment study, we used the time spent in front of an art piece as a proxy for visitor appreciation. We computed the aggregate time each piece was viewed and presented this total in a histogram organized by room and medium. The color of the bar indicates the medium whether it is a painting, a video, or a sculpture. The horizontal line in correspondence with each room shows the average

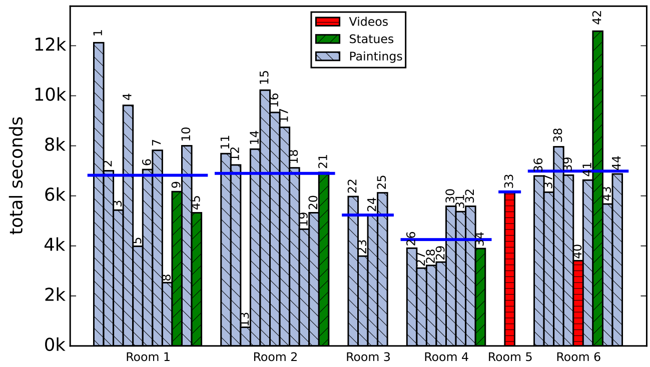


Fig. 5.6 Histogram showing the popularity of each artwork organized per room. Each bar represents the total number of seconds all the visitors have spent in front of each artwork. We also show the average time spent in each room, represented by the blue horizontal bar.

value across the bars (see Figure 5.6). The visualization shows that artwork `42` (i.e., a piece that included a live chameleon) was most popular followed by `1`, the artwork placed right at the entrance of the exhibition (it was also used as the poster for the exhibition around the city). Medium did not seem to predict popularity (for instance sculptures did not, on average, receive more or less attention than paintings). It is also interesting to note that the video documentary labeled as `33` received much more attention than the artistic loop video, of comparable length, labeled as `40`.

Looking at room-level popularity, one can notice two things. First, most of the time was spent in the first two rooms and in the last one, though visitors also allocated comparable time to room `5` due to the video documentary. Rooms `3` and `4` however received substantially less attention. These rooms were later in the exhibition and contained more abstract works. Second, the variance of time spent in these two rooms is relatively small, suggesting that no piece in these rooms attracted a great deal of attention. Rather, visitors seemed to walk sequentially through these rooms, without stopping at any piece in particular (in an example of the "museum shuffle"). This hypothesis is confirmed by the analytics we present next.

5.6.2 Visitors common movement paths

Once we know when a visitor is facing a particular piece and for how long, we can construct their path within the museum. Given we do

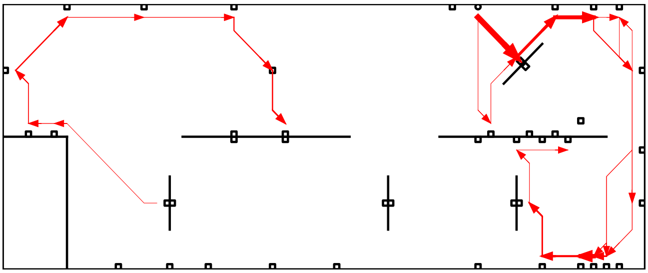


Fig. 5.7 Most frequent paths followed by the participants. The thickness of each line is proportional to the frequency of the represented subpaths.

not have absolute positioning information, we consider a path as the sequence of artworks that a person has faced during her visit, ordered by time. For our goal, this representation summarizes the path taken by the visitor without the noise contained in long sequences of absolute coordinates, and it is well suitable, as we will show, for data-mining tasks. When we project this sequence onto the actual layout of the space to visualize the path, we choose the shortest trajectory that connects the artworks and does not pass through walls. Hence, in our visualizations, a displayed path represents a logical trajectory in space. Also, for readability, we display paths by computing them over mesh previously described and displayed in Figure 5.2. In this way, we minimize clutter and present information in an orderly fashion.

To visualize routes, we constructed a visualization of the most popular paths. Based on interview data with a member of the van Gogh Museum staff, visualizations of all visitors together prove confusing. Given enough visitors, the image is easily filled with lines and all informational value is lost. Moreover, group behavior appears only indirectly, emerging from the overlap of the individual paths. Instead, we decided to take a direct approach that would minimize the effort of the user to recognize group behavior. The result is a visualization of the layout with a number of lines that represent common paths, whose thickness is proportional to their frequencies across all visitors. We proceeded as follows.

First, we computed which paths and subpaths were common among all the individuals. In particular, we computed the longest common subsequences (LCS) between all paths, and their frequencies across all paths. For example, suppose that two visitors at some point have faced artworks `1`, `2`, and `3`, in this order. In our statistics, we only considered subpath `1-2-3`, and not subpaths `1-2` and `2-3`, as they are included in LCS `1-2-3`. Note that two visitors can have multiple LCSs of different lengths, for example accounting for common subpaths in different rooms.

Second, we merged the 10 most frequent LCSs into a single visualization. Since LCSs can overlap, computing the thickness of each LCS requires to account for the frequencies of each overlapping subsequence. For example, consider two LCSs, namely `1-2-3` and `2-3-4` both with associated frequency 1. Merging the aforementioned LCSs requires to increase the frequency of the overlapping subpath `2-3`. In this example, the result is a final path `1-2-3-4` where subpaths `1-2` and `3-4` have an associated frequency of 1 while `2-3` has an associated frequency of 2. Hence, the line would be thicker in correspondence to the line connecting `2` to `3`.

Figure 5.7 shows the resulting visualization. One can notice a number of aspects. First, visitors tended to turn left after artwork `1`, and approached it as first piece or right after artwork `6`. While of scenic effect, the initial wall might have confused the visitors who did not know what was "expected" from them. Second, visitors tended to walk along the external edges of the space. In an independent observational study, this same pattern was observed in the Van Gogh Museum. Third, some artworks are not reached by the line. It is possible that some of those paintings were often skipped by the visitors, perhaps because recognized and ignored while walking the dominant external path, hence appearing less frequently in the paths. Another more likely explanation is that due to their position outside of the dominant line, they were reached by the visitors at different points of the path, hence generating less frequent LCSs. We verified this claim by visualizing the common paths generated by choosing the 20 most frequent LCSs.

It is interesting to notice how room `3` and `4` do not include the line at all. As mentioned earlier, according to our measurements these rooms contained less popular paintings, and were sometimes completely skipped. Also, the layout of the room requires a decision by the visitor in front of each divider wall, again whether to visit the right-side or the left-side of the room first. Similar behavior was observed in another study [136], when museum visitors entered spaces without an obvious point of orientation their paths scattered in different directions. It is likely that visitors took very different decisions, if they decided to visit the room, and visited the pieces in very different order, perhaps skipping some. Again, this would generate less frequent LCSs that would not make it to the top 10.

5.6.3 Common positions of visitors

In this study we track proximity to art pieces rather than absolute location. Although we lack this information, we can leverage the coordinates provided by the particle filter, and the fact that often a mobile point detects multiple anchor points at the same time, to compute a reasonable approximation. While through the enclosures we steer the detection ranges towards a face-to-face measurement, there are still areas of overlap between the ranges where a mobile point detects multiple anchor points at the same time. This is particularly true for pieces positioned very close to each other. In occasions when a mobile point detects multiple anchor points the particle filter will predict a position that is in, or close to, those areas of overlap. That is because indeed it is likely that a person was effectively positioned not precisely in front of an artwork. Moreover, over a certain window of time, it is more likely that the mobile device will detect the closer anchor more frequently than the further anchor, and this information can be exploited by the filter as well. In this sense, particle filters applied locally to a number of close-by anchor points act as a localized localization system, though we still lack distance measurements for a precise localization. In other words, particle filters compute a sort of multi-lateration between close-by anchor points, for example, to triangulate the position of the visitors in a specific area of the museum. Finally, we can also sometimes estimate positions further from anchors detection range. For example, if we measure a visitor leaving an exhibit and appearing at another exhibit further away after some short time, and this time is coherent with walking speed and the distance between the two exhibits, then we can position the visitor around the line connecting the two exhibits during that time.

We computed the estimations of the absolute positions for each visitor, for those moments in the visits where the estimate confidence was over a threshold we chose empirically. With these estimates, we produced the heatmap presented in Figure 5.8. The areas of the heatmap are colored according to the total amount of time visitors have spent in that area. While it does not precisely contain the same information, we can notice that this visualization resembles a spatial representation of the histogram presented in Figure 5.6. The visualization confirms that most of the visitors time was spent in room `1`, `2`, and `6`. Also, as artworks `13`, `14`, `15`, and `16` were within a distance of 1 meter from each other, the smoothing phase of the heatmap generation aggregates their values and it is difficult to distinguish artwork-related information. This is acceptable, and in fact expected, as the visualization shows spatial estimation.

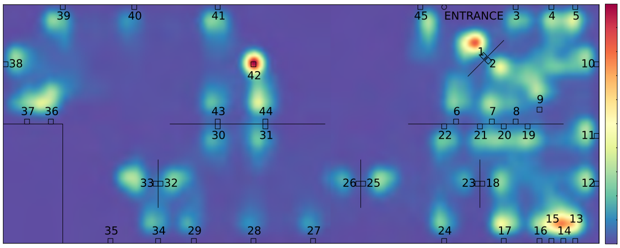


Fig. 5.8 Heatmap showing the amount of time spent in areas close to artworks. Warmer colors are associated with areas where the participants have cumulatively spent most time.

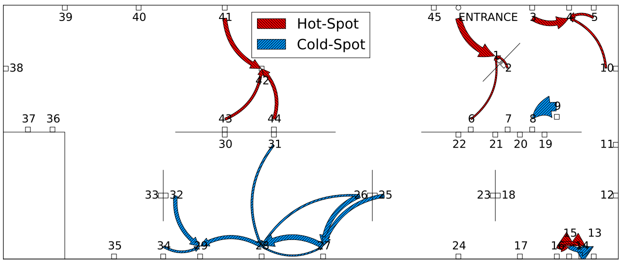


Fig. 5.9 The relationship between the most (red arrows, IDs 42, 1, 15, 4, and 16) and the least (blue arrows, IDs 13, 8, 27, 28, and 29) viewed artworks respectively and their neighboring pieces. The pictures show the neighboring sources covering 50% of the total paths.

5.6.4 Artworks attraction and exhibition design

In some exhibitions, curators and designers guide visitors through a logical sequence of artworks. Such a sequence could be intended to describe a story or to keep the visitor engaged and interested during her visit. However, visitors might follow a different path or be unexpectedly attracted by some pieces more than others. To provide an insight about how artworks attract visitors, for example highlights, we produced the visualization depicted in Figure 5.9. Such visualization is based on the following analytics: by selecting the five most (*hot-spots*) and the least (*cold-spots*) attractive artworks we show their relation with the pieces that commonly precede them during visits (i.e., paths). All the arrows represent the most common transitions from one piece to the targeted one and the thickness relates to the frequency of each transition for that piece. In other words, a thicker arrow indicates that many visitors have first visited the source piece and the target piece right after. Additionally, we limit the number of incoming arrows to the ones that cover 50%

of the total frequency for each target. In this manner, we focus our attention to the source artwork from which the target mostly attracts the participants.

Figure 5.9 shows information related to the capacity of hot- and cold-spots to attract visitors. On the one hand, piece `1` mostly attracts visitors from the entrance and less from the neighboring pieces. For artwork `42`, instead, visitor transitions are *evenly* distributed among the most immediate pieces. This result suggests that visitors may approach `42` due its attractiveness more than due to a sequential scanning of the room. However, such effect could also be attributed to its central placement with respect to its neighbors. On the other hand, cold-spots have less evenly distributed arrows, with a preference for the previous pieces. If we look at the artworks in room `4`, we observe that the attraction is catalyzed towards the left part of the room (while it appears so, it is not a display of a path). In fact, the thicker arrow consistently comes from the closest piece on the right. This suggests that those artworks were mostly viewed while passing through the room, further suggesting, together with their popularity measures, the lack of interest of the participants in these pieces.

5.6.5 Focus group on visualizations' effectiveness

To understand the utility of the visualizations, we conducted an initial study with the art director, the curator, and director of business development of CoBrA. Due to the limited size of this study, its outcome cannot be considered and generalized as a main result of this work, yet it is valuable to bootstrap a discussion about the strengths and limitations of our approach, and the opportunities for future work. We followed a "think aloud" method [37] in which respondents interpreted the visualizations for us as they reviewed them. We discussed the visualizations two times, first with minimal explanation from us of what is depicted, and then again after we provided more explanation. In both cases, we asked staff members to explain the content, and share their interpretation. We ended with questions on which visualizations were most helpful and a discussion about the overall impact of the system.

The respondents were able to decipher most of the visualizations. At some points, they emphasized a different aspect of a visualization. The most pronounced example of differences in interpretation were for the histogram presented in Figure 5.6. The artistic director and the curator focused on which pieces were most and least popular, for example referring to the popularity of the authors, while the business developer looked at the distribution of time across rooms. In others, their interpretations coincided. After looking at artworks popularity (Figures 5.6 and 5.8), visitors paths (Figure 5.7) and the transitions to hot spots (Figure 5.9), staff came to the conclusion that the route was clearer in the rooms in which they spent more time and that visitors were more likely to move unpredictably in the rooms in which they spent less time.

All respondents used the visualizations to reason about visitor behavior in both the show depicted and future ones. Although the business developer found the bar chart of time distribution accessible, the artistic director said that bar charts were not her "language", yet both respondents interpreted the visualization with ease. They used that visualization alone and in conjunction with the visualization of common paths to begin to question what they saw in the exhibition overall, why certain pieces and/or rooms commanded more attention. For example, noting that people spent less time and did not have a clear route in two rooms in the later half of the exhibition, they both wondered if later rooms commanded less attention because the work was less appealing or because of a dip in energy. The visualizations generated questions for them to test and they discussed possible solutions for the show under development including adding more benches later in the show and providing more guidance in those areas.

The respondents agreed on the value of some visualizations over others. They were most interested in paths through the museum (Figure 5.7) and the heatmap (Figure 5.8) showing time spent in different areas of the exhibition. They found a visualization of transitions to least popular pieces least relevant and transitions to most popular pieces somewhat relevant (Figure 5.9). The curator was in particular surprised in noticing how about 10% of the visitors visited the exhibition starting from the end. Respondents requested the data to share with staff in discussion of future shows and made only minor suggestions for improvements (i.e., adding the images of the art pieces to the visualization).

The visualizations altered how they thought about several issues. The visualizations raised "awareness" of the relationship between "content and structure" of the show. The artistic director acknowledged that most of her thoughts about design focused on the content and that these visualizations emphasized for her the importance of finding the appropriate structure for the upcoming shows. For all respondents, viewing the visualizations also highlighted questions for which there is no correct answer, e.g., what is the goal of design? Is it to lead the visitor a certain way or simply provide guidance? Although for them, the answer differed from show to show, the artistic director concluded that she wants people to “linger” and she wants to “allow freedom within structure”.

5.7 Clustering visitor behavior

In this section we present the application of data-mining techniques applied to visitor data. We used the set of r_v vectors and s_v path sequences from all 182 volunteers as dataset for two clustering tasks: (i) identifying common paths chosen by visitors when visiting the museum, and (ii) identifying patterns of the distribution of visiting time across rooms and exhibits. For both tasks, we utilize *Hierarchical Agglomerative Clustering* (HAC) [96], a bottom-up clustering algorithm where items initially form individual clusters and are merged in a series of iterations based on a clustering criterion. We chose the Ward method [145] as a criterion for choosing clusters to merge at each step, which focuses on minimizing total within-cluster variance.

The input of the algorithm is the distance matrix between all items in a dataset. To identify common paths, we compare all s_v sequences with the Jaro-Winkler (Jaro) [151] similarity metric, which is used to compute string similarity. Jaro is a type of string-edit distance that considers explicitly the mismatch in the order in which elements appear in two sequences (an operation called transposition) and how far mismatching elements are in the sequences. Intuitively, inverting in s_v the order of two exhibits that are nearby in the sequence is less penalized than inverting two exhibits far from each other in the sequence (e.g. $jaro(\text{“}ABCDE\text{”},\text{“}ABCED\text{”}) > jaro(\text{“}ABCDE\text{”},\text{“}AECDB\text{”})$). Precisely, we use $1 - jaro(a,b)$ as Jaro computes similarity while HAC requires distance. For the task of identifying patterns of time distribution, we compute the Euclidean distance between all r_v vectors. Before computing distances, we pre-process r_v vectors as follows. First, we use a threshold such that r_v contains only elements larger than 15 seconds (that is we consider for each visitor only the exhibits where she spent more than 15 seconds), and then we scale and center each adapted r_v. In other words, we transform r_v vectors into vectors describing how visitors distributed their time, among those exhibits where they spent more than 15 seconds. We then fed both datasets to the same clustering algorithm.

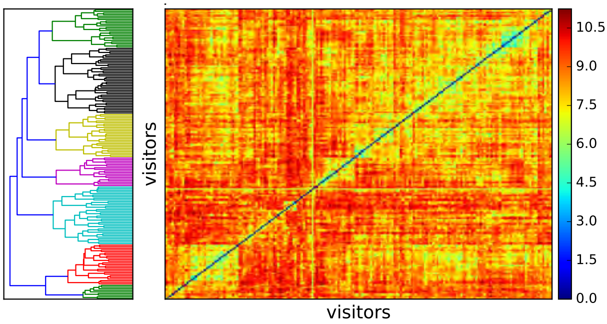


Fig. 5.10 Hierarchical agglomerative clustering of the visitor time distribution vectors. At the center the distance matrix shows the distance between the vectors of all the visitors, while the dendograms show how the visitors have been grouped together.

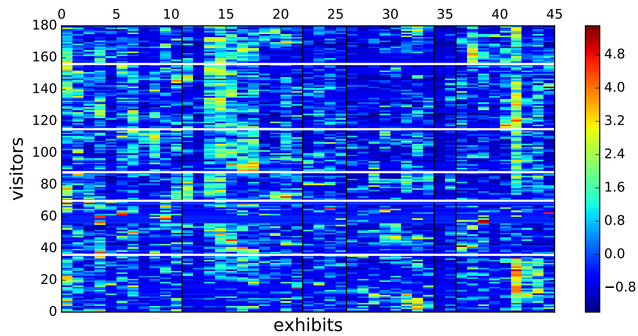


Fig. 5.11 The clustering algorithm identifies one small group (1-38, bottom) that spent time mostly in room 6, and another major group (top) that is further clustered. Horizontal white lines show cluster divisions and black vertical lines room divisions.

5.7.1 Time distribution across exhibits

In Figure 5.10 we show the distance matrix between the r_v vectors, which are organized according to the result of the agglomerative clustering displayed in the linkage dendograms, and in Figure 5.11 we show the set of pre-processed r_v vectors grouped by the result of the clustering algorithm (vertical black lines show room divisions and horizontal white lines show cluster division). The dendogram is a tree displaying the hierarchical clustering process, where each vertical line represents a merge between two clusters (the length of the horizontal line shows the distance of the merge). Because each cluster starts from a single row, one can notice that each merge starts by merging two rows together up to the root. The distance matrix shows the distance between each visitor. One can notice that distance between visitors in the same cluster is lower than between visitors between visitors belonging to different clusters. Because the visitors are organized according to the results of hierarchical clustering, the distance matrix can give an intuition of the boundaries of the clusters that are merged at each step.

The results show that the clustering algorithm identifies two major clusters. The first cluster includes the bottom visitors (1-38, light green and red in the dendogram) for their particular interest in 3 artworks in room `6` and little interest in room `2` and `3` except for exhibit `1`. The second cluster includes all visitors between (65-182) who spent time in room `2` and in front of exhibit `42`. This second cluster is further clustered in other clusters, for example, (i) (85-118, yellow) due to specific interest in the exhibits `7`, `8`, `9`, (ii) (124-160, black) due to specific interest in the first three exhibits, (iii) (160-182, green) due to some specific interest in exhibits `38`, `39`, `40` and `41`, as well as `19`, `20`, and `21`. Finally, HAC puts all outliers in a cluster together (38-65, light blue), though some of them do show a pattern, as it is the case of 40-50, showing a common interest in exhibits `2`, `3`, `4`, as well as `14`, `15`, and `38` and `39`.

5.7.2 Clustering of paths

We applied the same technique to paths. In Figure 5.12 and Figure 5.13 we show two of the clusters identified with HAC. We constructed the visualization as described previously, but focusing only on each chosen cluster, and this time selecting the most frequent 20 LCSs.

The first cluster includes the largest group of visitors and shows the most common behavior. HAC identified further groups within this major cluster, for example, splitting visitors turning right towards exhibit `6` from visitors turning left towards exhibit `3` after exhibit `1` at the entrance. The second cluster in Figure 5.13 shows a cluster with 10% of the visitors who visited the exhibition space in "inverted order", starting from room `6`. Perhaps these visitors did not understand what was expected from them by the curators. Similarly, through HAC we identified another group of visitors who decided to visit room `6` first, only to return right back to room `1` and continued from there, perhaps after realizing their mistake.

Furthermore, we can notice that both groups make more different choices of paths while visiting the first rooms, whereas the last rooms

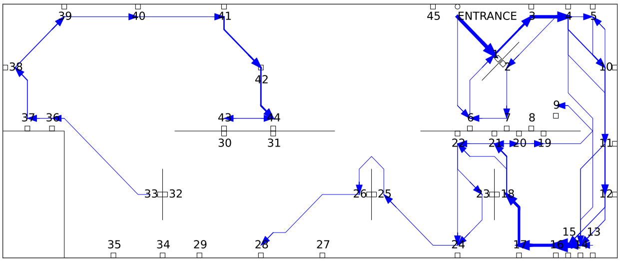


Fig. 5.12 Visualization of the common paths chosen among the largest group of visitors.

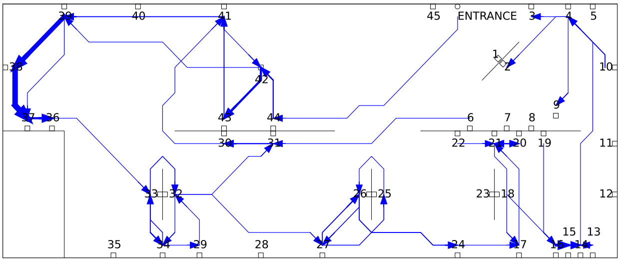


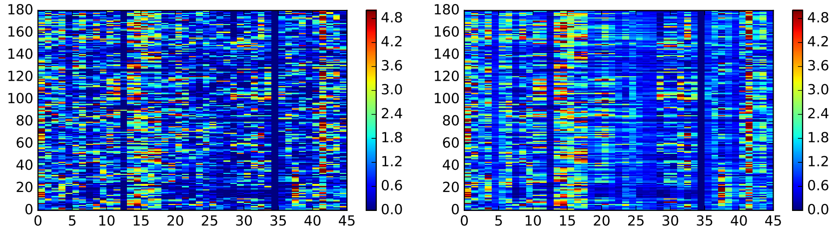
Fig. 5.13 Visualization of the common paths chosen by the 10% of visitors that visited the exhibition starting from the end.

are characterized by one common path where visitors scan the room sequentially along the outer walls. This phenomenon was dubbed "museum shuffle" by the staff of CoBrA and was associated to a decrease in attention after around 30 minutes in the visit.

5.8 Predicting visitor behavior

The previous section has shown patterns of group behavior in the data. That is, a group of individuals made similar choices during their respective visits and hence distributed the time in a similar way across artworks. Note that when we say that groups of individuals made similar choices, we do not mean that those visitors were together at the museum. In fact, most of the times they were not, and in fact visited the museum independently at a different time. Instead, the underlying assumption is that what these visitors shared was similar tastes and habits, which guided them to make similar choices.

If such patterns exist, we should be able to exploit them to predict visitor behavior. If a visitor presents similar patterns of time distribution for some artworks as other visitors (i.e., she belongs to a cluster), we should be able to leverage what we know about those visitors' visits



(a) Original rating matrix. (b) Reconstructed rating matrix.

Fig. 5.14 (a) The original rating matrix generated from the r_v vectors and (b) the reconstruction of such matrix from lower-rank matrices through Non-negative Matrix Factorization.

to predict the behavior of the visitor at other artworks. In other words, we should be able to apply the "people who liked this item also liked these items" approach notably used in recommender systems, and in particular in collaborative-filtering approaches. These approaches are usually applied to predict user-item ratings based on other ratings.

To test this hypothesis, we model our data as follows. We generate a 182×45 matrix R of ratings, where cell $R[i, j]$ models the rating that visitor i has given to exhibit j. In other words, similar to the matrix introduced in Section 5.7, each row is constructed starting from a r_v vector. For each row, we compute the ratings based on the amount of time spent by the visitor facing the exhibits as follows. First, we standardize each cell by removing the median across the row and scaling according to the Interquartile Range (IQR). The IQR is the range between the 1st quartile (25th quantile) and the 3rd quartile (75th quantile). This is similar to removing the mean and scaling according to the standard deviation (i.e., the standard technique), but it is more robust to skewed datasets, as in our case (i.e., each visitor has many exhibits with low or zero values). Standardization of a dataset is a common requirement for many estimators. Then, we scale each cell in the row linearly to the continuous interval $[0, 5]$, such that the cells with the highest value are mapped to the rating 5 and the cells with the smallest value (often the value zero) are mapped to the rating 0. It is important to proceed with these two steps per-row, as visits have different lengths. The choice of the interval $[0, 5]$ is arbitrary and was chosen to provide an intuitive understanding of the mapping and the scale of the errors. This procedure produces the matrix presented in Figure 5.14(a).

Once we have produced such matrix, we can use a state-of-the-art technique for ratings recommendation. We choose *Non-negative Matrix Factorization (NMF)* [167], as it is an established technique with an intuitive meaning[1]. The idea behind NMF is to compute two matrices whose product should produce the original matrix (i.e., R in our case). NMF assumes the presence of some latent variables in R that can be identified and exploited to re-compute the original matrix. The two matrices usually have a lower rank, corresponding to the number of latent variables expected in the data (a reasonable choice is usually around the number of clusters expected in the data). When lower-rank matrices are used, NMF often produces values larger than zero in cells corresponding to un-rated items. These values can be used as predictions. We choose a rank of 15 though we had comparable results in the 10-30 range. For reference, Figure 5.14(b) shows the matrix as reconstructed by NMF using the full R matrix as input. One can notice that the patterns are still present in the reconstructed matrix, with an additional smoothening of the isolated values.

To establish the prediction capability of our model, we apply NMF to R as follows. We divide the ratings in R between a *training set* and a *test set* by choosing uniformly at random r ratings of the whole set, with $r = 0.01, 0.05, 0.1, 0.2, 0.4, 0.8$ (hence we try to predict up to 80% of the original matrix using 20% of the data), and use these values as test set and the rest as training set. Then, we use the ratings in the training set to re-compute R, and compare the ratings in the test set between the values in the reconstructed matrix and in the original matrix. In practice, (i) we copy R, (ii) we set to zero the cells in the test set, (iii) we compute NMF of this matrix, and (iv) we compare the values in these cells to compute the prediction error, computed as the absolute difference between the two. Note that test set ratings can be initialized to the medium or mean of the corresponding row or column, but for simplicity we initialize them to zero (if we initialize the values to the mean of the corresponding column in the training set, we noted particularly an impact on larger values of r). We repeat the test for each value of r 100 times.

Figure 5.15 shows a box-and-whisker plot with the distribution of the prediction errors for each value of r. The box delimits the IQR, the band and the square in the box are the median and the mean respectively, the whiskers are set at the 1.5 of the IQR, and outer points are the outliers. We can see that the median prediction error is consistently beneath 1

[1] NMF gained general popularity in the field when a technique based on NMF won the Netflix prize.

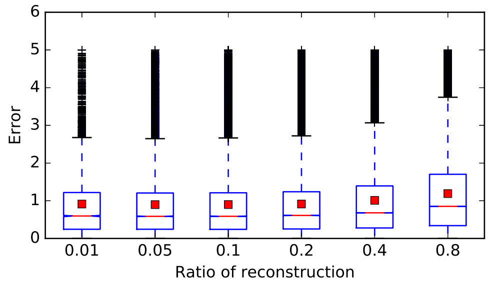


Fig. 5.15 Distribution of prediction error across depending on the ratio of reconstruction.

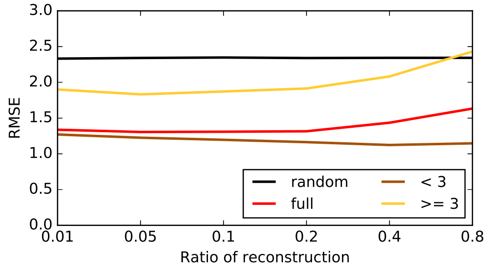


Fig. 5.16 RMSE when choosing the test set with different criteria and against a random predictor baseline.

for small values of r, increasing together with the error range only for large value of r. One thing to notice here is that outliers are also denser as r increases. These elements suggests that the error is accumulated around the higher ratings, while NMF is more accurate in predicting low ratings.

To validate this hypothesis, we compute the Root-mean-squared error (RMSE) aggregating all tests, selecting the test set in three different ways: (a) we select a test set across R as described previously, (b) we select a test set across R choosing only ratings smaller than 3, (c) we select a test set across R choosing only ratings larger than or equal to 3. In addition, we compute the error with a random predictor as a baseline. Figure 5.16 shows the RMSE for each value of r. One can notice that for small values of r NMF can achieve low RMSEs, that grow for r larger than 0.2. Moreover, we can also notice that errors are accumulated around large ratings more than by the smaller ones. These errors may be accumulated by those ratings that are outliers to the trends, as noticed previously by the smoothening effect of NMF in Figure 5.14(b). These results show that the data collected by the proximity sensors not only show group behavior, but that can also be used to predict the behavior of the visitors. Note that in terms of pure RMSE, we obtained lower errors when initializing the test set cells to the mean value of the column (i.e., around 1.6 for $r = 0.01$), but minimizing the RMSE is out of scope of this work.

5.9 Discussion

The proposed solution is able to reconstruct the behavior of a group of visitors with a pipeline of software filters. Through a simple and inexpensive sensing infrastructure, we were able to accurately position visitors at paintings regardless of data loss and noise caused by our enclosures and the absence of distance estimation from anchors.

We plan to investigate how estimating distance from anchors could impact the sensitivity of the measurement, in particular in areas with higher density of anchors where our technique produces more ambiguous positioning. Similarly, we would expect substituting our enclosures with tailor-made directional antennas to produce more robust and reliable communication ranges, and hence more accurate measurements.

With the widespread use of wearable sensors, like smart watches, glasses, and bracelets, pervasive and ubiquitous sensing capabilities will extend further than those provided by smart phones only. A fusion of different data sources should allow us to better quantify the quality of the proximity relationships between visitors and artworks.

While the visualizations show that certain artworks were more popular than others and that visitors consistently spent less time in certain rooms, the curator found that popularity is not necessarily the only metric to be trying to optimize for. The curator takes into account many different factors when she decides which artworks to exhibit, depending on the type of exhibition. Certain exhibitions often mix popular with less popular but historically or contextually related artworks, while so-called "blockbuster" exhibitions may indeed target more popular works of famous artists. Moreover, both the curator and the exhibition designer found that it is often expected that visitors focus only on certain pieces and may ignore some rooms completely. Nonetheless, it is difficult to justify the consistent ignoring of certain rooms by the vast majority of the visitors.

In general, it is not obvious how to translate visitor behavioral data into a metric of success or satisfaction regarding the exhibition. While there is a correlation between the amount of time spent at an exhibition or in front of an exhibit and the reaction of the visitor, as shown also by our results in behavior prediction, the quality of that reaction needs to be further explored. The curator found that having behavior data available to the museum staff without a further understanding of such metrics could lead to misinterpretations about the outcome of the decisions of some members of the staff. The internal response of the visitor could be investigated with other types of sensors in conjunction to proximity data as, for example, physiological sensors [78] and accelerometers [99] (see Chapter 7).

Finally, related to topic of understanding is the problem of how to intervene to improve or steer the response of the visitors, once issues have been identified. The strategies and heuristics of intervention remain a realm of the expert personal experiences skills of the museum staff, but further research is necessary, perhaps based on tools such as those presented in this paper, where differences in visitor behavior are measured as a result of curatorial changes (which would require longer (longitudinal) studies).

5.10 Conclusions

We presented a method to measure person-to-object relationships via face-to-face proximity. The positionings extracted from our sensors can be processed through data mining techniques that identify group behavior. The approach is inexpensive and requires little setup. The method provides a reliable measurement of positioning that is not bounded to the particular technology used to sense face-to-face proximity, and can hence be used in a variety of scenarios and applications. Our approach is not specific to our devices or to a particular medium, and it applies to any technology that can map onto the proposed binary model. Furthermore, results suggest the ability to use low cost sensors to capture visitor behavior metrics of interest to staff and the qualitative results indicate the immediate benefits of the use of these analytics in practice. Future work is needed to explore more conclusively the benefits of a visitor analytics tool for museum staff.

6 Exploiting density to track human behavior in crowded environments

> *The best time to go to Disney World, if you want to avoid huge crowds, is 1962.*
>
> — **Dave Barry**

6.1 Introduction

In the previous chapter we have discussed how mobile and fixed proximity sensors deployed in a museum can help museum staff understand the behavior of their visitors. However, the work of museum staff of popular museums does not end with curatorial decisions. It also includes, for example, making sure that visitors can approach all exhibits and information without congestions and clogging, that they have access to all resources such as restaurants, restrooms and lockers when needed, and that they do not experience queues and waits that are too long. Ensuring a safe and fulfilling experience is an important goal of the work of museum staff that is particularly challenging when the museum is crowded. Staff operating popular and crowded museums need to monitor visitor density continuously, to intervene when necessary and to support planning and decision-making. However, as crowd density increases, sensors are known to provide more noisy, incomplete, and ambiguous data, problems that are exacerbated by complex indoor venues. In other words, museum staff are left with instruments that operate unreliably in those conditions where they are most needed.

Part of this chapter have been originally published in **"Exploiting density to track human behavior in crowded environments"** C. Martella, M. Cattani, M. van Steen - *IEEE Communications Magazine* 2017, and have been slightly modified to improve readability.

Contributions. In this chapter, we present a MObile-Nodes-Assisted positioning system (MONA) that can operate at conditions of high crowd density, by utilizing proximity sensing between mobile sensor nodes (i.e., sensors worn by the visitors) as well as between mobile and fixed sensors nodes (i.e., so-called anchors). Traditional approaches that rely only on mobile-to-anchor sensing, fail when density increases and mobile sensors cannot sense anchors. This is mainly due to the shielding effect of the body of the visitors, which attenuates the radio signals and increases loss of detections. Instead, our approach exploits the increased presence of mobile sensor nodes in the surrounding of each individual, effectively transforming a challenge into an opportunity. In fact, positioning accuracy *increases* when more visitors wearing a sensor are present in the museum. Through a multi-hop approach, and by leveraging radio signal strength information, MONA can position a visitor (i) when mobile-to-anchor proximity detections are missing and even (ii) at points of interest not instrumented with anchors.

The remainder of this chapter is organized as follows. First, we provide an overview of the museum, the requirements of the staff, and the sensing infrastructure. Then, we present the related work on museum technology and positioning systems. After we describe MONA and our complete filtering pipeline, we evaluate the deployment of the system in a real-world museum. In particular, we show how the data can be used to understand the behavior of visitors during the different events take place in the museum. Finally, we conclude by discussing limitations and future work.

6.2 Overview

In this section we provide an overview of the museum where we deployed MONA, the requirements of the museum staff, the sensing infrastructure, and the data analysis pipeline.

6.2.1 Museum

The museum subject of our deployment is the 5th most visited museum in its country, with half a million visitors per year and more than three thousand daily visitors during our experiment (we chose the three most crowded days, just before Christmas time). With a median visit duration of three hours, the museum can present extremely crowded scenarios with peaks of nearly three thousand people visiting at the same time.

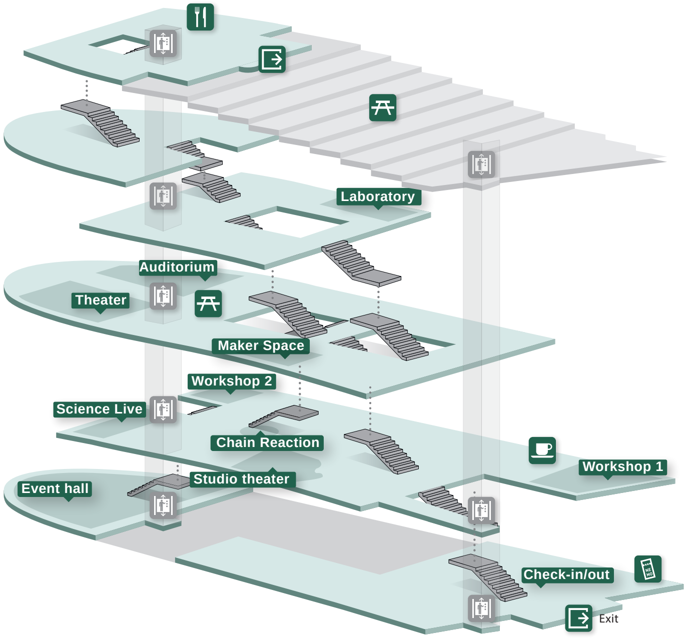


Fig. 6.1 The map of the museum. The layout of the museum presents challenging aspects, such as multiple stairs and elevators connecting the floors, an open-space nature as well as the presence of closed rooms (e.g., auditoriums, labs, and theaters) and a series of balconies projecting over the underlying floors.

Approximately 25% of all visitors participated in our experiment, with a peak of around 600 participants at a single moment in time.

Unlike other museums with clear paths to follow and different well-divided rooms, our museum mostly consists of an open multi-story area, deployed over six stories and with virtually no boundaries. The first four stories share a large hall in the middle connected by several stairs, with floors 3 and 4 that are effectively balconies projected over the underlying stories. With this layout, network links spanning over multiple floors are common. Figure 6.1 shows the 3D map of the museum.

Since the museum targets mainly children and families, the observed mobility patterns present lots of intense and fuzzy dynamics. The museum is not only an unstructured open space but it also offers different types of attractions: events scheduled in closed rooms (workshops), events scheduled in open areas (shows) and unscheduled attractions all over the place (interactive exhibits). Exhibits can comprise multiple

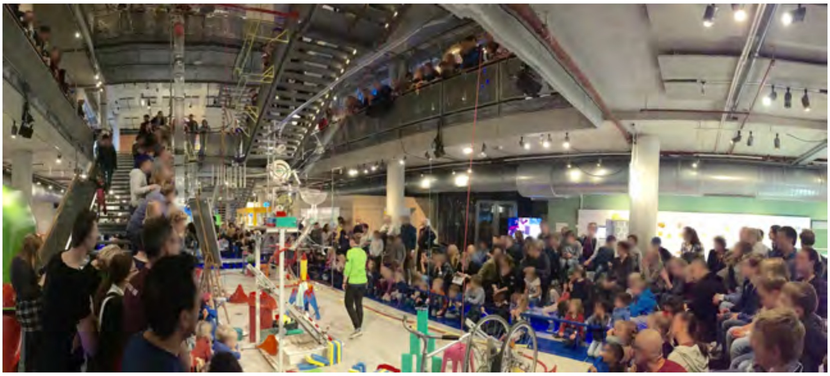


Fig. 6.2 An example of the Chain Reaction event that periodically attracts dense crowds, sometimes from different floors. One can notice individuals around the event location, as well as on the upper floors.

items within a radius of approximatively 3-4 meters. Finally, the museum offers two snack bars at the first and second floor and a bigger restaurant at the top floor. Among the scheduled attractions, the most popular is a show called “Chain Reaction” which spans from the first to the third floor of the building. This show is repeated 5 times a day, lasts 15 minutes and attracts a good part of the visitors on the internal balcony of the museum. Figure 6.2 shows a snapshot of the Chain Reaction show taking place. As this event influences the behavior and flow of the visitors and can get quite crowded, we dedicate part of our data analysis later in the chapter to this attraction.

Overall, the museum curators were interested in three types of information: (i) the popularity of a set of points of interest and exhibits, (ii) the distribution of visitors across the floors, and (iii) the flows of people between floors. Based on these requirements, we decided to deploy our positioning infrastructure to collect data and answer their questions.

6.2.2 Sensing infrastructure

To collect positioning information, visitors were asked to wear a *bracelet* equipped with a 2.4GHz transceiver and a microcontroller running a neighbor discovery protocol with a sampling rate of 1 Hz. The bracelets were powered by a super capacitor, reducing the device charge time to under 2 minutes. The lifetime of the bracelets was 8 hours with a duty cycle of approximately 2%. This allowed us to charge and program the bracelets just before a check-in procedure, where visitors were informed of the experiment and decided whether to participate or not (no personal information was collected and all data was anonymized). The check-in/out procedure additionally created a unique entry in the database with a timestamp and the bracelet ID, that allowed us to remove any exchanged messages between the anchor points and the bracelets that had yet to be handed out.

Anchors. For each point of interest (PoI), an *anchor* device was installed. These devices had the same hardware as the bracelets but were different in two ways. First, they were running a modified version of our neighbor discovery mechanism that did not require anchors to discover their neighborhood, but instead only to be discovered by the bracelets worn by the visitors. Second, anchors were powered by either a wall socket or by a large battery back. Through a reliable source of power, it was possible to run the neighbor discovery algorithm with longer duty cycles, increasing even more the probability of a PoI to be discovered by bracelets. As a result we were able to discover bracelets in the range of a PoI in just a few seconds. Note that while for positioning purposes it is usually more reliable to instrument each PoI with an anchor, as we will show for `MONA` this is not a necessary condition.

Every second, bracelets could receive broadcast from anchors or other bracelets within a distance of some 5-7 meters. Such broadcasts contain the unique identifier (ID) of the sender, and we consider the reception of such broadcast as a *proximity detection* between the two nodes involved. Bracelets recorded together with the ID of the other node also signal strength information about the broadcast (which can be used as an approximation of the nodes' distance).

Sniffers. Once every second, bracelets reported the proximity detections collected during the previous second via a special packet sent on a dedicated channel to the backbone of the system consisting of the *sniffers*. Like anchors, sniffers used the same hardware as bracelets in addition to a single-board computer that used either WiFi or Ethernet to commit the packet to a central database hosted on a server. The sniffers were placed uniformly in the museum to cover all areas, with some overlap.

Each sniffed packet could report up to three detections, favoring two anchors and one bracelet when possible. The resulting dataset was collected over three days of experiments containing more than 17 million anchor-to-bracelet and bracelet-to-bracelet detections, together with timestamp and signal strength information. In this chapter, we focus on the second day of experimentation.

6.2.3 Data analysis

Once data is collected by the sniffer and committed to the central database, it can be analyzed either in real-time or offline. Because of the mentioned irregularities of wireless communications, in particular at high density, the database presents a number of shortcomings, such as missed detections and detections that are logically unexpected when considering location and signal strength. We designed a filtering pipeline to overcome these artifacts and compute the positioning of the visitors at PoIs at each time of their visit. The pipeline comprises two steps.

First, the set of proximity detections is filtered with a particle filter that takes into account the topology of the museum, the positions of anchors and PoIs, and the position and movements of the visitors wearing a bracelet. This step drastically increases positioning accuracy compared to more naive approaches, such as assigning visitors to the anchors detected with highest signal strength, but also allows positionings at PoIs that are not instrumented with an anchor (which is impossible with the aforementioned naive approaches).

Second, we smoothen the positioning information with a density-based filter and a majority-voting filter, which remove some of the artifacts and oscillations left over by the particle filter. At the end of the pipeline, we are able to assign at each moment in time each visitor to a PoI, if any. We utilize this filtered dataset of positionings for further analysis about PoIs popularity and flows between floors.

6.3 Related work

There is a large body of literature regarding indoor localization and positioning with mobile sensors. In this section we first present existing work regarding use of pervasive sensing technology in museums, and then we describe existing efforts and limitations of indoor (cooperative) localizations systems.

Museum monitoring. In the case of monitoring and analyzing the behavior of visitors in a museum, the visualization of metrics such as popularity, attraction, holding power and flows has been explored to support the work of museum staff [85, 133]. Earlier works focus on localizing visitors at coarse-grained levels (room level) through technologies like Bluetooth [162] to support multimedia guides [22, 149]. Nevertheless, more complex sensors and localization techniques can be employed to obtain a deeper understanding of crowd related phenomena.

A recent work, for example, utilizes "physiological" sensors and a positioning system based on RFID to classify the behavior of visitors into three categories: "the contemplative", "the enthusiast", and "the social experience" [78]. Similarly, the location and heading of pairs of visitors have been used to classify the visitor behavior early in their visit and to provide social-aware services to increase their engagement [42]. Differently from these approaches, our work does not focus on a classification task but it tries to quantify the behavior of visitors throughout the whole museum area. Moreover, our work extends existing work enabling measurements in scenarios where the sequence and boundaries of exhibits are not clearly defined.

An alternative approach to crowd monitoring with mobile and wearable sensors is to use cameras. Computer-vision techniques are a common approach to tracking human mobility [166, 130, 161] and detecting anomalies in a crowd [103, 81, 148]. However, cameras can suffer from poor lighting, and temporary or permanent obstructions [166]. Also, fusing the views from multiple cameras together in a highly dynamic indoor scenario is quite challenging [128]. Finally, the privacy issue of collecting large-scale footage of visitors often restricts researchers from accessing these sources of data (as in our case).

Indoor localization. A simple way of monitoring a crowd consists in precisely tracking the absolute position of each individual. While this technique is a feasible option in outdoor scenarios – where the GPS system can be exploited – for indoor conditions accurate localization is still an open problem.

Among the wide literature of radio-based localization techniques, only few [118, 14, 90] are accurate enough to be employed in museum scenarios. Unfortunately, none of these techniques perform consistently throughout the museum areas. It is known, in fact, that the localization error increases significantly at the edges of rooms and in hallways [94], conditions that are often present in museum topologies. For our tracking application, this phenomenon can be even more problematic, since even small estimation errors could lead to visitors being associated to the wrong exhibit, positioned in the wrong room or placed on the wrong floor.

An additional problem of indoor localization techniques is that the location of a device is often estimated by comparing the current radio observations with a fingerprint obtained at deployment time. Unfortunately, in crowd monitoring applications, the ever-changing conditions of the museum and the crowd can drastically affect the network connectivity, undermining the efficacy of fingerprint comparisons. Changes in the furniture location, for example, could lead to different multi-path effects and, thus, different radio fingerprints [94] and hence require recomputing fingerprints of the environment. Moreover, radio characteristics such as the received signal strength can be significantly influenced by the density of the crowd, up to the point that this relation is exploited by some mechanisms to estimate the crowd density [163, 105].

We borrow some ideas from the field of cooperative localization [150, 157, 108]. In cooperative localization, nodes cooperate and share measurement information to increase their localization accuracy in a peer-to-peer manner. Typical information being shared between nodes is the geographical topology and some measurements regarding the communication between nodes with known and unknown locations. In contrast, we are not interested in deploying algorithms that can be computed in a distributed manner by the nodes "in-the-network". We focus instead on computing bracelet positions through a dedicated central repository, leveraging a more global and complete view without being limited by network topology. This aspect is particularly useful in high density scenarios, where packet loss increases, potentially hindering cooperative approaches.

Finally, many of these approaches are usually tested outside of the extreme conditions of high mobility and density of a complex real-world museum as the one subject to our study.

6.4 Model

We consider N visitors $V = \{v_1, v_2, \ldots, v_N\}$, each wearing a bracelet during the visit and being uniquely identifiable during the T seconds of duration of the day (in our case, from 10AM to 6PM). While bracelets have their unique IDs and were re-used during the experiment, each element in V has a unique ID assigned at check-in, and we use these IDs for detections. The museum has O points of interest $I = \{i_1, i_2, \ldots, i_O\}$ and M anchors $A = \{a_1, a_2, \ldots, a_M\}$. In the simple case where each anchor is assigned to a PoI, $I \equiv A$. We consider S the set of proximity sensors $S = V \cup A = \{s_1, s_2, \ldots, s_{N+M}\}$, as the union of detectable anchors and

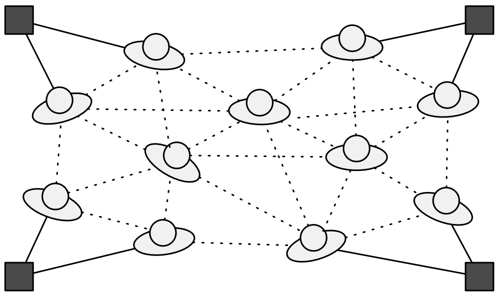


Fig. 6.3 The problem can be thought as unfolding a lattice of nodes with unknown location, leveraging the known locations of the anchors and the signal strength of mobile-to-mobile and mobile-to-anchor detections.

visitors. We define the series of proximity detections for a visitor v as a $S \times T$ matrix D_v, where $D_v(i,j) = r$ if the proximity sensor of visitor v detected sensor s_i at time j with signal strength r. Note that D_v has $N+M$ rows, as it comprises all proximity sensors, either used as anchors or worn by other visitors. $D_v(*,t)$ refers to all detections collected at any time t, and $D_v(i,*)$ to all detections of sensor s_i. Moreover, we define a positioning matrix M_v as the $O \times T$ matrix where $M_v(i,j) = 1$ if the visitor was at a distance shorter than d (i.e., the detection range, in our case 5-7 meters) from PoI i at time j. Note that there can be times where a visitor is not positioned at any PoI, and that a visitor cannot be positioned at more than one PoI at the same time (when a visitor is within d distance from multiple PoIs, we choose the closest).

6.4.1 Problem definition

Our goal is to produce the positioning matrix M_v starting from the series of detections in D_v. In principle, in the case of $I \equiv A$ and with perfectly working sensors (i.e., with signal strength correctly mapping to distance and no inter-floor detections) we could use D_v to generate directly M_v by assigning a visitor to the anchor detected with highest signal strength, and computed to be at distance shorted than d. As the visitor walks around the museum, we would generate contiguous series of "bits" in M_v reflecting the intervals of proximity with the anchors and PoIs. However, real-world scenarios like ours present the following shortcomings.

First, often proximity sensors miss detections from sensors close-by within detection range. Missed detections can be caused by shielding effects of the body of the visitors, as well as the environmental interferences. When proximity detections are lost, it is difficult to position the visitor at PoIs. Second, proximity sensors can collect ambiguous detections, for example, from anchors spatially close-by but at the other side of a wall or at a different floor. Note that these proximity detections are often received at a high signal strength, so they can generate wrong positionings. Moreover, due to tunneling effects, spurious proximity detections can also happen, though more rarely, between sensors further than the expected transmission range (tens of meters). Third, signal strength is unreliable and noisy as a proxy for distance estimation, as it can vary given the same distance between visitors and anchors depending, for example, on the crowd density as well as the museum topology.

Finally, in real-world scenarios it is not always possible to instrument every PoI with a corresponding anchor node. Anchors are instead spread over the museum (for positioning purposes) and installed in convenient locations, without being associated to a particular PoI. In this case, it is possible to position visitors at un-instrumented PoIs by exploiting the detections of close-by sensors, which can be both fixed (other anchors) or mobile (other visitors), and whose sensing range overlaps over the un-instrumented PoI.

In principle, the problem can in fact be thought as the unfolding a lattice of nodes with unknown locations, connected by edges with weight proportional to the strength of the respective mobile-to-mobile detections, by leveraging the known location of the anchors and the strength of mobile-to-anchor detections[1].

Note that this aspect of `MONA` has commonalities with some indoor localization techniques, but has the simpler goal of assigning a visitor to the closest PoI within distance d, and not necessarily a precise absolute position (although it may require to estimate one as a step to achieve such goal). To a certain extent, our task can be thought as computing a more coarse-grained location within the specific boundaries of the anchors in the local areas of the museum.

6.4.2 Particle filter

We designed and implemented a particle filter that takes into account the topology of the museum, the location of anchors and PoIs, and the estimated location and movement of the visitors. Particle filters have been successfully used in indoor localization to estimate the absolute position of individuals with unreliable sensors [56, 63]. For localization,

[1] In this sense, the problem is also analogous to the embedding of a graph through a string-based force-model.

usually a mobile sensor communicates with a few anchors installed at known locations. It is assumed that the sensor can communicate with all, or a majority of, anchors from all positions and directions, and that the sensor can measure distance from these anchors, for example, through signal strength or time-of-flight. Our setup is more complex, as we have a larger number of anchors that are detectable at shorter distance, and a number of detections from proximity sensors at *unknown* locations, i.e., the visitors, together with a multi-story venue.

The filter requires topology information about the museum, such as the sets of anchors A and PoIs I, each defined by a triple f,x,y with f being the floor number and x,y the coordinate within that floor, and a set of walls $W = \{w_1, w_2, \ldots, w_M\}$, each defined as a segment between two points. Every floor is modeled as a distinct two-dimensional space, with an independent origin. As floors are connected by a number of different stairs and elevators, it is difficult to model the transition spaces between floors reliably. Instead, as we discuss later in this section, we assume a visitor can "appear" at a certain floor and model this transition confidence in the filter, based on the measurement. For each visitor we define a set of particles $P = \{p_1, p_2, \ldots, p_K\}$, each defined by a tuple f,x,y and a weight w that models the likelihood of the visitor to be at that coordinate. Initially, particles are spread uniformly at random across the museum floors.

For each time of the day $0 \leq t < T$, the following four steps are executed for each visitor checked-in at that time, given the respective detection matrix D_v.

- **Estimation**: At each time, we estimate the floor the visitor is positioned at, if any, by computing the floor with the largest number of particles (normalized by floor sizes), given enough confidence is provided by the particles. Then, we compute the likelihood of each particle's estimate (i.e., its position) given the measurement at time t, that is the set of detections in D_v contained in the t-th column. For each particle p, the weight is updated using the likelihood function $\Phi(p, D_v(*,t))$. More details about this step are given later in this section.
- **Positioning**: We estimate the position of v by computing the weighted average among the floor's particles (i.e., *the centroid*) and find the closest PoI i_i on the floor within distance d. We then set $M_v(i,t) = 1$, unless the confidence of the estimate is smaller than a threshold δ.

- **Re-sampling**: We create a new set of particles by drawing with replacement from the current weighted set of particles. While drawing particles from the set, we favor particles proportionally to their weight (i.e., their likelihood). As a result, particles with higher likelihood are picked more often than particles with lower likelihood. Depending on the confidence in the current set of particles, we may choose to (i) pick only from the particles in the current floor, (ii) additionally spread particles with smaller likelihood across other floors, or (iii) re-distribute all particles across all floors (i.e., when we believe we have lost track of the visitor). More details about this step is given later in this section.
- **Movement**: We move particles at walking speed in random directions, avoiding illegal moves, such as walking through walls.

Compared, for example, to previous work on positioning [100] (see Chapter 5), the two steps where we focused most of our efforts to tailor the filter for our use-case are the estimation and the re-sampling steps. Those steps were redesigned to (i) include mobile-to-mobile detections, (ii) consider multi-floor environments, (iii) support PoIs without anchors and anchors not associated to PoI, and (iv) omni-directional sensing with signal strength.

In general, we can compute the confidence of the particles' centroid by measuring the dispersion of the particles and the amount of time spent without collecting any detection from a sensor located on the same floor as the visitor. We consider the visitor to be at the floor with the largest number of particles, normalized by floor sizes, given enough confidence. Note that having a visitor assigned to a floor does not mean we position the visitor at any PoI, as that depends on the position of the particles within the floor, their likelihood, and dispersion.

Likelihood function $\Phi(p, D_v(*,t))$ computes the likelihood of particle p given the set of sensors detected at time t that were positioned *on the same floor* as p. In Φ we consider both anchors, of which we know the location, and neighbor visitors, for which we use the centroid computed at time $t-1$. In other words, we use neighbor visitors as anchors. We consider neighbor visitors only if we are confident enough of their centroid.

Intuitively, the likelihood of a particle depends on the particle's distance from a reference point proportionally to the strength of the detections. This reference point can be a single sensor or an average across sensors. We compute the reference point and the distance from it as follows. We first map signal strength values to the $(0,1]$ continuous interval by means of a Gaussian kernel[2]. Then, if multiple sensors were detected, we compute the average coordinate across these sensors' coordinates weighted by the respective signal strength, and use the particle's distance from this average coordinate. We use a linear kernel to map distances to the $(0,1]$ continuous interval.

If only one sensor was detected, instead, we compute the difference between the signal strength value and the particle distance from the sensor (both mapped to the $(0,1]$ continuous interval through the respective kernels), and use this difference. Note that while single-sensor reports are more common, the system is still practically multi-laterating over time, as particles that are more coherent with a series of measurements receive a higher likelihood.

When we update particle weights and re-sample the particles, we proceed depending on the centroid confidence c as follows (the two thresholds δ and δ' to be chosen empirically).

- $c > \delta$: This means we know the visitor is on the floor. We ignore the likelihood of particles from other floors, practically "teleporting" these particles from other floors to the current one.
- $\delta' < c \leq \delta$: This means we are starting to believe the visitor may have left the floor. We spread particles with smaller likelihood from the current floor to other floors.
- $c \leq \delta'$: This means we have lost the visitor completely. We re-distribute particles uniformly at random across floors.

As a visitor moves around a floor, particles spread towards areas that are more likely to have produced the current measurement. As the visitor moves to a new floor, absence of detections on the previous floor causes particles to disperse until we start spreading particles on the other floors, including the new one.

6.4.3 Smoothening

The particle filter is able to drastically increase positioning accuracy. Yet, some artifacts, such as short jumps to a PoI nearby or on another floor, are still possible though rare. These jumps produce gaps in the bursts of positionings as well as spurious positionings. We apply the remaining two smoothening steps of our positioning pipeline comprising a density-based filter and a majority-voting filter.

First, we apply a density-based filter [98] (see Chapter 4) to all the rows of M_v. The filter computes statistics about positioning information

[2] signal strength decreases non-linearly with distance

at different PoIs over time, and clusters them together. These clusters of positionings are effectively intervals of time where the visitor was in proximity to a PoI. Once these intervals are identified, the filter fills the gaps in these intervals, and ignores the noisy positionings. One can think of this filter as a low-pass filter implemented with a sliding window, but the optimal parameters of the low-pass filter are chosen automatically.

Then, we pass a majority-voting filter on the data produced by the previous stage. The filter slides a window of L seconds of duration ahead of, and including, t over all rows to decide at which PoI to position v. The filter decides by choosing the PoI with the largest number of positionings in that window of time, and breaks ties by choosing the PoI chosen at the previous window (or randomly if none was chosen at that time).

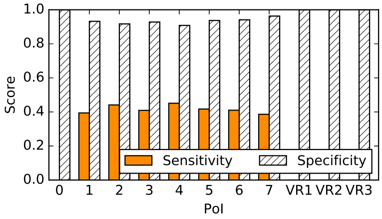
6.5 Evaluation

6.5.1 Methodology

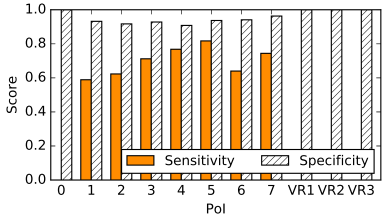
Before the 3-days main experiment, we conducted a *controlled* experiment where we scripted a visit of the first floor while the complete sensing infrastructure in the museum was turned on. The script defined arrival and departure times at each PoI. Originally, the first floor had all PoIs associated with an anchor, and one additional anchor to improve positioning accuracy. To test the ability to position visitors at PoIs without anchors, we added to the script three "virtual" PoIs not associated with any particular exhibit or anchor (that is, solely defined by a coordinate in space). In addition, one PoI with an anchor was *not* used in the script, but the anchor was turned on and hence could affect accuracy.

Together with the scripted visit, we distributed 15 bracelets around each interested PoI area within a radius of some 15-20 meters from the respective anchor (or virtual PoI coordinate), at 1 meter of height, and moved them at random across the space for the whole duration of the stop. This setup allowed us to control bracelet density and movement, and at the same time minimize external factors like body shielding effects (which were tested during the real-world experiment). Each stop lasted eight minutes, divided in four slots of two minutes. During the first slot only the visitor's bracelet was on, while during the following 3 slots we turned on the additional bracelets, in groups of 5 per slot.

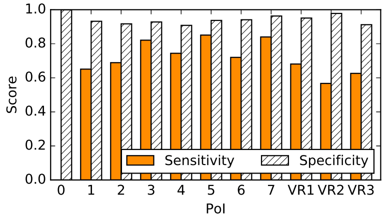
To evaluate real-world accuracy, for the duration of the main experiment we positioned 5 bracelets at known locations at PoIs, and additionally scripted a visit of the whole museum of the duration of around one and a half hours with 14 stops of the duration of around



(a) Strongest RSSI



(b) Smoothened Strongest RSSI



(c) `MONA` (anchors only)

Fig. 6.4 Positioning accuracy during the controlled experiment at the PoIs on the first floor and three additional "virtual" PoIs not instrumented with an anchor. We compare our technique (not using here mobile-to-mobile proximity) to a technique that positions the visitor at the anchor with strongest signal with and without a rolling window. The combined balanced accuracy, defined as the arithmetic mean between average sensitivity and specificity, for the two competing techniques and `MONA` (anchors only) were 0.62, 0.69, and 0.81 respectively.

5 minutes each (hence, not all time was spent at PoIs). Also for this scripted visit, we added a number of "virtual" PoIs such that the visitor did not stop at PoIs associated with an anchor, except for the stops at the restaurant and the Chain Reaction.

Setups. We implemented the filtering pipeline with less than 1000 lines of Python code. By implementing the filters with vectorized SIMD operations and leveraging linear algebra libraries like numpy[3] and a just-in-time compiler with numba[4], our pipeline can manage in real-time (i.e., sub-second compute time to process 1 second worth of data) peak-time data with a single thread executed on a ultra-low power laptop[5].

[3]http://www.numpy.org/

[4]http://numba.pydata.org/

[5]Provided with a mobile 1.2Ghz Intel Core M CPU

The pipeline can be easily ported to multi-cores and even GPUs for very large crowds.

For both experiments, we set filter parameters to the same values, that is 1000 particles per visitor, majority-voting window length of 10 seconds, and density-based filter ε and *minPts* to 10 and 3 respectively.

Metrics. To measure the performance of our solution at the task of positioning visitors at the PoIs we compute the number of:

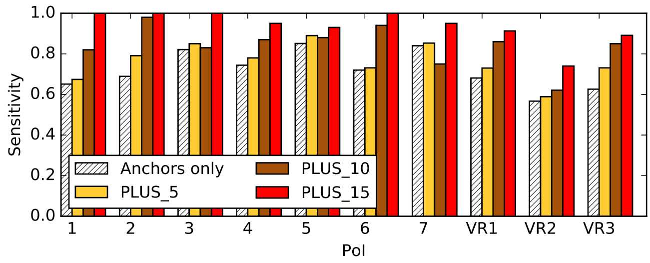
- *False positives* (FP): positionings that are present in the measurement but not in the ground truth.
- *False negatives* (FN): positionings that are present in the ground truth but not in the measurement.
- *True positives* (TP): positionings that are present in the measurement and in the ground truth.
- *True negatives* (TN): positionings that are missing in the measurement and in the ground truth.

We compute these values by comparing the positionings produced by the pipeline and the ground truth. We use these tests to compute two statistical measures of performance for binary classification tests. Sensitivity can be used to measure the ability of the test to identify positive results and is defined as $sensitivity = \frac{TP}{TP+FN}$. Specificity can be used to measure the ability of the test to identify negative results and is defined as $specificity = \frac{TN}{TN+FP}$. Intuitively, they measure the ability to correctly estimate positioning and its absence, respectively.

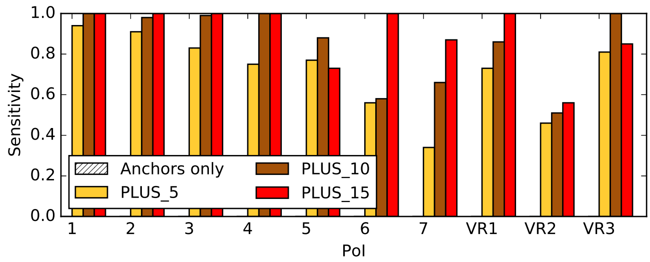
6.5.2 Results

Controlled experiment. We test our technique against an approach where we position the visitor at the anchor/PoI detected with strongest signal strength. As this technique is subject to missed detections and noise, we additionally extend it by making positioning decisions over a sliding window of 10 seconds. Precisely, we position the visitor at the anchor/PoI that presents more frequently in the window detections with largest signal strength (i.e., a form of majority-voting). Figure 6.4 presents the sensitivity and specificity of each anchor/PoI for the two techniques as well as for `MONA`. Note that for this test, we compute values for `MONA` only for the first slot, that is we position the visitor *solely based on anchor detections*.

One can notice that raw data suffers from missed detections, yielding an average sensitivity that is around 40% the sensitivity obtained when smoothening the decisions over a window (0.42 and 0.68). As expected,



(a) MONA



(b) MONA (ignoring anchors)

Fig. 6.5 Impact of neighbor bracelets around a visitor on the first floor. We added 5 more bracelets in the surrounding of the visitor every 2 minutes of each stop.

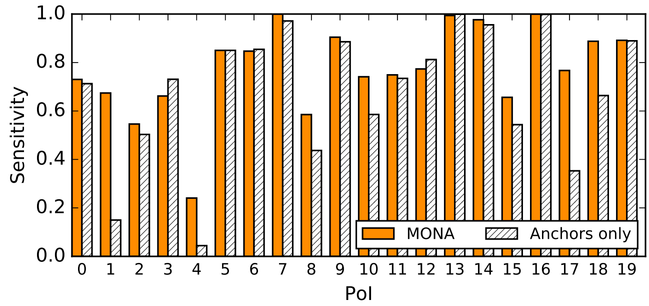


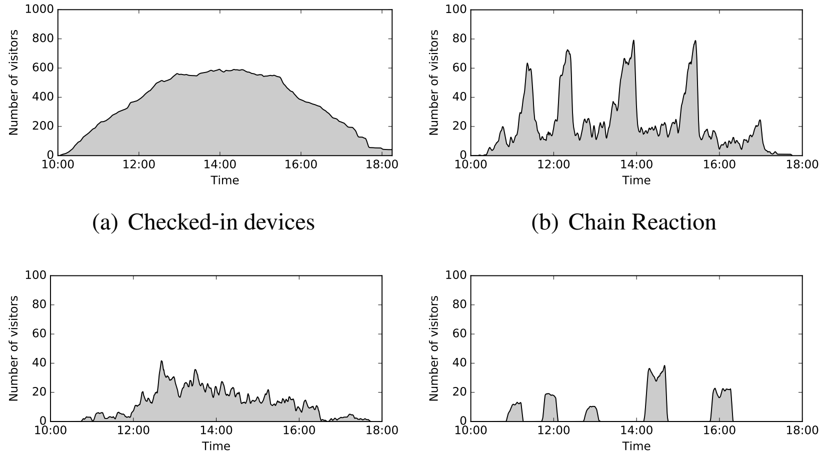
Fig. 6.6 Positionings accuracy of our system with MONA and without it during the real-world experiment. Only PoIs 7, 9, 13, 14, 16, 19 were instrumented with an anchor. The average sensitivity was 0.68 and 0.77 respectively.

specificity is close to the maximum value of 1, as it is difficult to wrongly position a visitor at a PoI far away with a controlled transmission range. Looking at the results of MONA (anchors only) one can notice a relative improvement in average sensitivity of over 5% (from 0.68 std. 0.34 to 0.72 std. 0.26), compared to the smoothened technique, but the most interesting result is the impact on positioning at virtual PoIs (impossible with the other techniques). This is due to the spatial nature of the filter.

Figure 6.5(a) presents the sensitivity results when `MONA` takes into account also detections of neighbor bracelets to position both the visitor and the neighbor bracelets. We do not present specificity values here as they are consistently close to 1 as in Figure 6.4. One can notice that adding mobile node proximity information improves positioning ability reaching a value of (or close to) 1 when 15 additional neighbors are used, with an average relative improvement between using only anchors and including 15 neighbors of around 30% (from average sensitivity of 0.72 std. 0.09 to average sensitivity of 0.94 std. 0.08). The improvement is slightly worse in the case of virtual PoIs (in particular VR2), but note that here we leveraged the anchors used for the other PoIs (except for only an additional anchor). Moreover, more data points (i.e., visits) should yield more statistically consistent results (e.g., for PoI 7 and 3 adding 10 bracelets yields worse results than using anchors only). Finally, one would expect a slightly lower impact in the real world, where body shielding effects and other irregularities would influence mobile-to-mobile detections.

To show the impact of mobile-to-mobile proximities, we compute the same results but *ignoring anchor detections completely* for the scripted visit. That is, we position the visitor *only based on mobile-to-mobile detections*, but position the neighbor bracelets based on both mobile-to-mobile and mobile-to-anchor detections. Clearly, this setup could not be deployed and is used only for testing purposes. Figure 6.5(b) presents the results of this test.

First, all values for the first slots are 0, as no anchor detections are available. Second, for certain PoIs the results are better than for Figure 6.5(a). This result may seam counterintuitive at first, as we are using less information (i.e., the anchors detections). However, the reason is to be found in the "averaging" effect performed by the neighbors. In fact, the positioning of the visitor cannot be affected by noisy detections from anchors far away but it is based only on the neighbors that are consistently positioned by the filter around the area (and those few that temporarily are not, are "filtered out" by those that are). Third, for some PoIs the performance is much worse, again in particular VR2. In general, these results shows how local mobile-to-mobile proximity detections can overcome missed detections from anchors, which tend to increase when crowd density increases.



(a) Checked-in devices (b) Chain Reaction

(c) Restaurant (d) Auditorium

Fig. 6.7 Number of individuals positioned at different PoIs over time. (a) The number of visitors wearing a bracelet during the day (experiment participation was around 25%). (b) The Chain Reaction PoI. Events started at 11:15AM, 12:15PM, 2:45PM, 3:15PM and 4:45PM for the duration of about 15 minutes. (c) The restaurant at the top floor. (d) An Auditorium open to the public only in 5 occasions during the day.

Real-world experiment. Figure 6.6 presents a comparison of sensitivity values, for both the stops of the scripted visit and for the 5 bracelets installed at PoIs, when mobile-to-mobile detections are used and ignored for positioning. Note that 13 out of 19 PoIs are virtual, hence they represent the hardest task for `MONA` (i.e., only PoIs 7, 9, 13, 14, 16, 19 represent stops at actual PoIs instrumented with a dedicated anchor). One can notice that `MONA` produces a relative improvement in average sensitivity of over 13% (from 0.68 to 0.77), though for a few virtual PoIs the performance is lower. Again, more data points should produce more consistent results.

A main reason why we do not always see a strong impact as presented in Figure 6.5(a) is due to the “sampling” effect of the real-world experiment. Considering that “only” 25% of the visitors were wearing a bracelet, the chances to leverage mobile-to-mobile proximity detections are reduced. This is aggravated by that fact that in the first and last part of the day crowd density is lower, as fewer visitors are in the museum. Lastly, at each second bracelet often report only one neighbor visitor, as bracelets favor reporting two anchors out of three detections when possible. We would expect a higher impact by increasing the number of reported neighbors. The current value of three was due to limitations to the packet size of the current implementation and can be increased in future versions.

6.6 Application

We now proceed with a number of analyses of positioning data regarding visitor behavior in the museum.

6.6.1 PoIs "popularity" over-time

Because at each time we know which visitors are positioned at which PoI, we can estimate how many individuals are visiting a PoI. In Figure 6.7(a) we present the number of checked-in devices during the day. One can notice that peak-time is between 1PM and 3:30PM, with around 600 visitors wearing a bracelet (and around 2400 individuals overall in the museum). Figure 6.7(b) shows the number of individuals spending at least two minutes at the Chain Reaction (CR) PoI throughout the day (hence filtering out individuals just passing by). CR events took place at 11:15AM, 12:15PM, 2:45PM, 3:15PM and 4:45PM for the duration of about 15 minutes. One can notice that CR events have the typical footprints of crowded events. First, in the *build-up phase*, density gradually increases during the minutes before the event, as people either stop-by or approach the event location in advance. Then, the event takes place and density remains more or less constant. Finally, in the *break-up phase*, density decreases quicker, as all individuals leave the event location for another PoI. We return to the study of this behavior later in this section.

Figure 6.7(c) shows data for the restaurant at the top floor. For this data, we aggregate positioning information for all the anchors at that floor, as the restaurant covers the whole space. Here, one can notice that lunch time peaks between 12:30PM and 1PM, but decreases gradually, as it is used by families for breaks at the end of the visit, enjoying the view of the city from the roof-top. Figure 6.7(d) presents data for the Auditorium. The Auditorium is a closed theater space, that is open to the public only during the scheduled events. For this reason, one can notice no visitors outside of scheduled shows and a less gradual build-up phase.

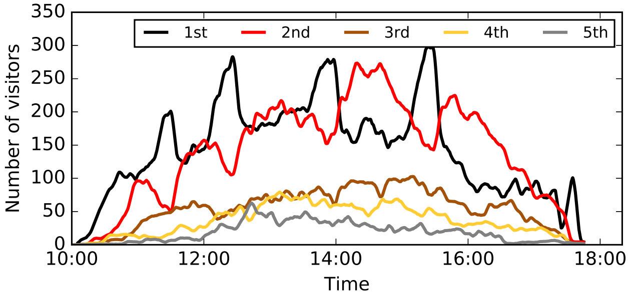


Fig. 6.8 Distribution of visitors across floors over time. As time passes, the relative "popularity" of upper floors increases, with visitors moving up the building during their visit. Also, the Chain Reaction event plays a big role in the way people move across floors, attracting visitors from other floors.

6.6.2 Crowd distribution across floors

Furthermore, we can look at how visitors are distributed across floors. Figure 6.8 shows the number of individuals at each floor throughout the day. First, one can notice how the CR event greatly influences the distribution of visitors across the first two floors and also at the third floor. When the CR event takes place, we can see corresponding peaks at the first floor and dips at second floor (and smaller dips even at third).

Second, one can notice that the "popularity" peak of floors happen at different times of the day depending on the floor. The first floor hosts more people during the first half of the day, while second floor peaks around 2PM, and third floor around 3PM (and the fourth floor has very few visitors at all before 12PM). This is because the building is built to be visited somehow in order floor after floor, though visitors are not obliged to do so. Again, the fifth floor hosts the restaurant and has a different pattern, dominated by lunch time habits.

6.6.3 Flows between floors

Finally, we can look at visitors flow from floor to floor. Figure 6.9 shows the number of visitors transitioning per minute *from* and *to* the first floor in particular. We compute this by considering only visitors that remain on the floor for at least 10 minutes (hence filtering out visitors just passing by). One can notice again that CR events, in gray, dominate the patter. Small peaks in the transitions to first floor appear at the minutes before CR events, while taller and more sudden peaks appear in the transition from the first floor right after the event (again, the footprints of build-up and break-up phases). As stairs to the second

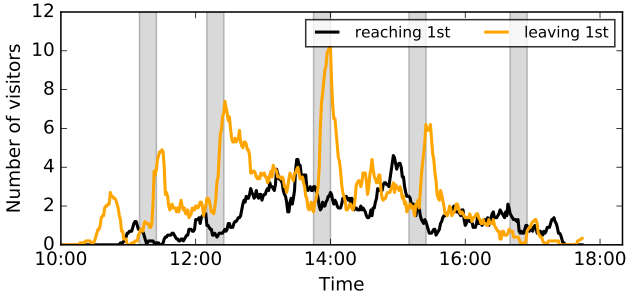


Fig. 6.9 Flow of visitors *from* and *to* the first floor. One can notice the influence of the Chain Reaction, as visitors tend to leave the floor by the stairs close-by the attraction *after* the event, and approach the floor from the other floors *before* the event.

floor are positioned right besides the CR event location, visitors tend to move to the second floor right after the event finishes.

6.7 Discussion

Regarding positioning accuracy, `MONA` uses mobile-to-mobile detections not only to fight density, but to leverage it and yield better results. Moreover, `MONA` can position visitors also at PoIs not instrumented with anchors. To improve this second task, one can increase the number of anchors around the museum, but it is an open question how to choose such locations in the most effective way. Increasing *anchor density* locally to specific areas, perhaps those hosting more exhibits and attracting higher visitor densities, should be a valid heuristic, but more research is necessary to understand what factors can influence such decision.

Being in close proximity to an attraction does not necessarily mean the visitor is paying attention to that particular attraction. Compared to spaces hosting paintings and other visual arts, the attractions of the museum were interactive and invited visitors to move around. Visitors were not required to face a particular direction, as in the case of a canvas on a wall. It can hence sometimes be erroneous to infer popularity just by counting the number of individuals at a certain PoI. For example, we noticed that PoIs close-by the CR event location present a small increase in the number of positionings when most crowded CR events are taking place. This could be caused by individuals that needed to stand further from the CR event location due to a large audience, inside the proximity range of those PoIs, It is difficult to design a sensor that is able to measure proximity in the complete surrounding of a visitor, as it is necessary in museums like the one where we ran our study, and at the same time face-to-face proximity, as required by other types of attractions and analyses.

The data we presented shows clear behavioral trends, yet it is not always obvious how museum staff can turn such insights into action. The museum staff re-organizes and re-furbishes the space every day, changing event schedules and locations as needed and depending on expected types and numbers of individuals. There are hence plenty of opportunities for the staff to make more data-driven decisions. For example, too large audiences at the CR event could be avoided by organizing other events in different locations at the same time, a strategy used by crowd management teams at music festivals.

A topic related to intervention is prediction. A number of different models exist to synthesize and predict the behavior of a crowd on the topic of evacuation, pedestrian mobility, and crowd dynamics [148]. While some of these topics are of interest to museum staff, such as those related to the safety of the visitors, none of them includes the particular nuances that characterize the entertaining and educational experience of a museum. Ad hoc models need to be developed and fed with sensor data, to better understand and predict how visitors react to the decisions of the museum staff. To this end, it may be necessary to fuse proximity data with data coming from other sensors, like accelerometers [99] (see Chapter 7) and physiological sensors [78], that can offer insights about different dimensions of the internal response of the visitors.

Finally, the experience of a museum does not start and end necessary inside the building. Long queues and congestions outside the museum are of great importance to the overall experience of a visit as well as to safety. Algorithms to predict and detect specific crowd dynamics such as pedestrian flows, queues, or clogging outside or at the periphery of the museum are still missing, and would be relevant to the more general field of crowd management as well [97] (see Chapter 3).

6.8 Conclusions

We have presented the design and evaluation of a positioning system that leverages mobile-to-mobile proximity sensing to overcome missed mobile-to-anchor detections due to high crowd density. We have shown that our approach is able to increase positioning accuracy when more visitors wearing a proximity sensors are present in the museum. The museum where we conducted the study presented extreme conditions of density and challenging conditions due to a complex multi-story open space. We have tackled these challenges by tailoring a filtering pipeline to our use-case. Yet, the approach is general and applicable to any proximity sensing technology that is able to detect neighbor sensors with and an estimate of distance.

Moreover, we have used the data to gain insights about the behavior of the visitors during the experimentation days. The insights show a clear behavioral trend, opening new questions regarding how museum staff can integrate such an approach in their work. The work we presented is however not limited solely to the use-case of a museum, but it is applicable to the general problem of indoors crowd monitoring.

7 How was it? Exploiting smart phone sensing to measure implicit audience responses to live performances

It's not dangerous that computers start to think like men, but that men start to think like computers.

— **Sydney J. Harris**

7.1 Introduction

There is more to the behavior of a crowd than what can be observed from the external. Monitoring and understanding the internal response of a crowd to its environment, which includes other members of the crowd, any event happening in the surroundings, the location etc., can help practitioners implement practices that increase the comfort and safety of the crowd. For example, it can help the managers of a train station schedule trains at platforms in a way that not only minimizes congestions and density, but also distress. Moreover, it can help the organizers of festivals and parades, events where high density is a standard condition also in the most favorable cases (e.g., near a stage or in connecting roads), monitor whether such conditions are source of frustration, or worse of more extreme moods and emotions like panic.

An instrument providing information about the experience of an event by a crowd opens a number of different applications that go beyond crowd management. For example, the information about the experience of an audience of a performance, and how it changes over time, could

The contents of this chapter have been originally published in **"How was it? Exploiting smart phone sensing to measure implicit audience responses to live performances"** C. Martella, E. Gedik, L. Cabrera-Quiros, G. Englebienne, H. Hung - *ACM Multimedia* 2015, and have been slightly modified to improve readability.

help the designers of such performance gain feedback to improve their production. Moreover, individual responses to performances could help organizers, like theaters and museums, target individual customers with marketing campaigns more effectively (e.g., recommend similar acts when a positive response is detected).

Detecting the internal experience and response of a crowd through spatio-temporal information is difficult, and perhaps proximity sensors alone are not well suited to the task. Accelerometers, galvanic skin response sensors, and microphones are examples of pervasive sensors that have been explored in the past, and we focus on the former as they are currently installed in most smart phones and hence suitable to large-scale pervasive measurements.

As we show in this chapter, it may not be sufficient to focus on the accelerometric data of a single individual to understand her response, but it is necessary to look at the data of the group of individuals she belongs to as well. We need to understand how the whole crowd responses to make sense of the individual experience of its members. This may be necessary to filter out irrelevant and noisy movements, and identify instead the salient moments where the movements have stronger meaning (i.e., in terms of internal response). Proximity information can play a role in this particular aspect, as we need to characterize the group of people, and the part of the crowd, the individual is embedded in to perform such analysis.

We also show that response can be quantified by looking at the moments before and after an event, which are often characterized by mingling and networking activities (at least in social contexts). We show that after an even individuals can move differently during conversation, again requiring to detect conversations and f-formations with spatio-temporal information in the first place.

Let us consider the example of the Chain Reaction event presented in the previous chapter. We have shown that spatio-temporal information enables the identification of when crowded events take place and their duration, as well as the number of individuals involved. Furthermore, we have shown that it is possible to recognize the typical fingerprints of the build-up and break-up phases, and how these influence density and flows across different floors. Yet, we are unaware of how individuals have experienced the events. We cannot analyze the data of all the visitors all together, but we need to detect first the groups in the crowd that are attending different events, or are visiting different parts of location, to then perform higher-level behavioral and response analyses on these subgroups.

Contributions. More concretely, in this chapter make the following contributions. We show that (i) even when people are sitting and watching a live dance performance, they spontaneously react it via body movements that can be captured from a standard acceleration sensor, (ii) moments of common spontaneous bodily reaction correspond to memorable events in the performance as reported by survey responses relating to the performance, (iii) their reactions can be used to predict their enjoyment of the performance, whether they felt immersed in the experience, would recommend it to others, or thought dance performance changed their mood positively, (iv) and finally by considering the social context that surrounds the activity of going to a live dance performance, we also provide initial results using acceleration and proximity sensors, that suggest that a change in the mood of a person as a result of watching a live dance performance is reflected in their general body behavior while mingling.

7.2 Related work

Traditional methods to investigate the response of an audience to a live dance performance make use of self-reports, such as surveys and interviews [21, 113]. Digital technologies can overcome the limitations of surveys and interviews, by giving more direct and fine-grained insights into the response of an audience. For example, the explosion in popularity of the social media, e.g., Twitter, and mobile computing has broadened the borders of a live performance, as fans comment and post information and opinions live to the online community [16]. Practitioners are interested in the activity of their audience through the social media to understand both their response and to leverage their activities as marketers of their performances [69, 88]. For example, some theaters, including Broadway, have experimented with so-called "tweet-seats" reserved for customers who promised to tweet about the performance live [137].

Other rather less pervasive technologies can also overcome the granularity issues of surveys using sensors. For example, work in neuroaesthetics use fMRI scanning to relate viewer responses to the aesthetics of the performance [20, 38, 24]. Moreover, tracking of eye gaze from video has been used when trying to distinguish novice from expert observers of dance [131]. Finally, physiological sensing such as galvanic skin response (GSR) sensors, have been investigated to measure the arousal of individuals watching a video of a dance performance, and its relationship with the individuals' self-reports [86]. Similarly, GSRs have been used also to measure the response to other types of live performance, such as comedy [144] and movies in a cinema [58].

These attempts show an increasing interest in quantifying the experience of arts and cultural events, such as live dance performances. Unlike these approaches, we advocate the use of pervasive sensors which are readily available in smart phones, which enable less obtrusive measurements and on a massive scale, compared to those obtained via physiological sensing. In particular, in this work, we focus on using acceleration and proximity sensors to measure people's reactions to live performance, which have been used thus far to measure very different phenomena.

Specifically, most work that consider accelerometers and people have addressed the problem of activity recognition of daily activities such as walking, running, sitting, climbing the stairs [83], daily household activities including eating or drinking, vacuuming or scrubbing, lying down [12], or identifying modes of transportation taken [115]. There is a trend moving towards the detection of medically relevant events, such as fall detection [46, 168], but all of these approaches focus resolutely on physical activities where the behavior can be represented directly by quite specific movements of the body. It is possible to classify these types of activities with excellent performance, yet these activities are very different to analyzing the response to a live performance. Few works do exist where less specific body movements have been classified. For example, Matic et al. also used acceleration to detect speaking status by strapping an accelerometer to the chest so that vibrations directly caused by speaking could be detected [102], Hung et al. [71] used body movements to predict socially relevant actions with a device hung loosely round the neck or for detecting conversations[72]. Such works highlight the potential of measuring spontaneous bodily responses to external stimuli using more pervasive sensing.

Apart from focusing on different activities and tasks, the above mentioned works measure behavior in environments that are far less challenging than a theater where the audience sits in silence, and where the link between activity and behavior is not as direct. The most similar work to our own was presented by Englebienne and Hung [53] who found that they were able to identify audience members as professors and non-professors from their behavior while attending an inaugural lecture. Although they were sitting, the small movements made in reaction to parts of the lecture demonstrated implicit responses of interest to particular moments and content delivered during the lecture. However, they did not analyse whether reactions from the audience to the lecture correlated with enjoyment of the lecture, for example. Another closely related work where the audience response was measured was presented by Bao et al. [13] who investigated how users watching movies on a buffer tablet could have their implicit responses sensed by a wide variety of modalities from the tablet itself including the video, audio, tablet interactions, and accelerometer. In this case, movements from the tablet whilst the user was holding it were used to gauge responses. Using a multimodal approach they were able to predict the rating of users to movies they watched on the tablet. However, in this case, the user sat alone to watch the movies and was not inhibited by the social norms usually adhered to in an auditorium.

7.3 Case study 1: direct responses to a performance

To inspect whether it is possible to predict responses to a performance by using data collected with wearable sensors, we have conducted an experiment in an actual dance performance. We start this section by explaining the characteristics of the performance and the resulting dataset we obtained. Then, we describe the features we used in the data analysis and classification experiments. The next two sections present the qualitative analysis and classification experiments we did based on questionnaire responses. The qualitative analysis aims to show that the movements of the audience are informative of their reactions to the performance. We do this by showing that the captured activity of the participants and common responses correspond to the salient events happening in the performance. The classification experiments present our methodology for automatically predicting a participant's evaluation of the event. The performance and further analysis presented show that the proposed method is indeed promising and worth investigating further.

7.3.1 Data collection

The sensor set-up. We organized a data collection experiment during a dance performance. This event consisted of almost an hour and a half of performance without intermission. It mainly consisted of dancing but also included monologues by the performers in Italian, while the music was mainly based on live cello arrangements but also included pre-recorded songs.

Using triaxial acceleration and IR cameras (for additional data verification), we recorded 41 participants watching the performance. The accelerometers were located in a custom-made device hung around every participant's neck. These devices recorded at 20Hz and were synchronized to a global time obtained by communicating through a wireless network. However, due to hardware malfunctions, only 32 accelerometers recorded data.

Furthermore, each device was equipped with a wireless radio, which we used to broadcast the device's unique identifier (ID) every second up to a distance of some 2-3 meters. The reception of such broadcast by the devices nearby can be considered a proximity detection. Since the device lies on the front of the chest, the radio transmission is shielded by the body, hence restricting the proximity sensing mostly towards the front of the individual. The device logs each detection on the on-board storage along with their timestamps. We used an energy-efficient MAC protocol [142] to allow the devices to communicate their IDs and detect each other's proximity.

In addition, the performance was recorded using a GoPro Hero +3 to analyze salient moments (i.e., favorite moments that were reported by the participants).

Survey responses. To evaluate the experience, a questionnaire was filled in by all 41 participants after the performance. Each questionnaire had 12 questions, where each group of three questions aimed to measure one aspect of the experience. The four aspects were "enjoyment", "recommendation (to a friend)", "immersion" and "mood changes". Each participant evaluated these aspects of the performance on a ten-point Likert scale, where one means "I completely disagree" and ten means "I completely agree". For measuring enjoyment, we adapted and selected questions presented in [124]. For the task of immersion, we selected involvement questions from the group Presence Questionnaire [123]. For recommendation we used items from O'Brien's questionnaire [106]. Each of these questions were carefully chosen to measure each task and slightly adapted to match our scenario. We formed the questions regarding mood by ourselves. Given that the majority of the audience members were Dutch, we used a back-translation procedure to ensure that each questionnaire item was accurately describing the original English wording. This involved finding three different Dutch speakers to translate the questions from English to Dutch, then from Dutch to English and then from English to Dutch again ensuring that the finally chosen words best matched the original English. From the total number of participants, 32 responded with the Dutch questionnaire and 9 to the English one.

Of the 32 participants with working accelerations, 25 reported a favorite moment of the performance. Two moments were particularly memorable: the *motorcycle sequence* was declared as favorite by 32% of the participants, and the *bolero finale*, favorite of 52% of the subjects. Note that in some cases that participants declared more than one favorite moment.

7.3.2 Feature extraction

We used the variance of the accelerometer readings, which is expected to act as a proxy for the physical activity level of the participants. Our assumption was that both subtle as well as more expansive movements of the the participants is related to the experience of the event. We expect participants with different evaluations of the event to have different movement patterns throughout the event, especially during salient parts of the performance. We calculate the variance in a sliding window of 2 seconds with 1 second shift, which corresponds to 40 samples for each window with a shift of 20 samples. This window size is carefully selected to capture the subtle and short variations in motion while still preserving a fine time scale and is empirically proven to perform well.

We extract features from an interval of $\sim$79 minutes, starting just before the first piece, when all participants are seated and ending when the final piece of the performance finishes. With this we obtain 4705 different variance values for each axis. Before calculating the variance along each axis, each axis is normalized by computing the z-score to remove interpersonal differences. We also calculate the variance of the magnitude, resulting in 4 different variance values for each interval. For the qualitative analysis, we only use the variance of the magnitude. For classification, we treat the variance values of each axis as well as the magnitude as our features, resulting in a 18820 dimensional feature vector for each participant. This feature choice restricts the representation to be temporally dependent. Therefore we may not capture cases where two participants liked (or disliked) the event during different parts of the performance. With another formulation of the problem, like using bag-of-words or a multiple instance learning approach, temporal dependence can be avoided. Although such cases are quite probable, in this

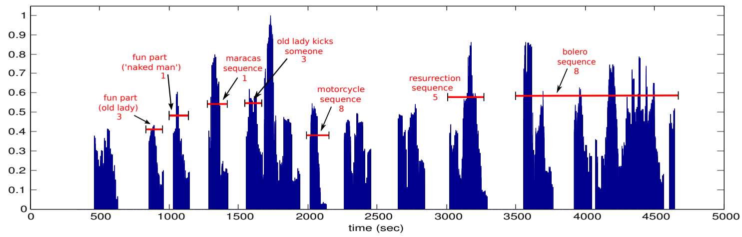


Fig. 7.1 Mean co-occurrence measurement distance over time for all participants using Mutual Information (MI). Salient moments are highlighted in red with number of appearance

experiment we assume that participants with similar evaluations of the performance tend to respond similarly during salient parts of it.

7.3.3 Data analysis

The variance in magnitude signals from all the participants were compared against each other to create a pairwise co-occurrence measurement over time using Mutual Information (MI). These signals were calculated over a sliding window (size of 60 samples) shifted by one sample, resulting in a vector of co-occurrence over time between two participants. The mean mutual information at each time interval was calculated, allowing us to evaluate the collective response of the participants to the performance over time. We hypothesized that salient moments would correspond to a high MI among all participants. These moments were chosen using a Otsu threshold [62] on the values of the computed signal. Figure 7.1 shows these salient moments captured by points where the mutual information goes beyond the threshold (blue), as well as the frequency of the reported favorite moments (red) that appeared in the free text survey responses. The time stamps were generated by manually identifying the time period(s) where the reported favorite moments occurred. Notice that the two moments declared as favorites for the majority of participants (*motorcycle* and *bolero finale*) are captured. This shows that memorable moments for people during these events can be captured by their coordinated movements, as they share the experience.

Furthermore, the role of music during the performance is also interesting and we want to understand its effect, if any, on the collective behavior of the participants. To do so, we looked at the sound intensity of the performance as obtained from the video and annotated song changes. Figure 7.2 shows the sound intensity of the performance (green) compared to the normalized co-occurrence measurement for MI (blue). The

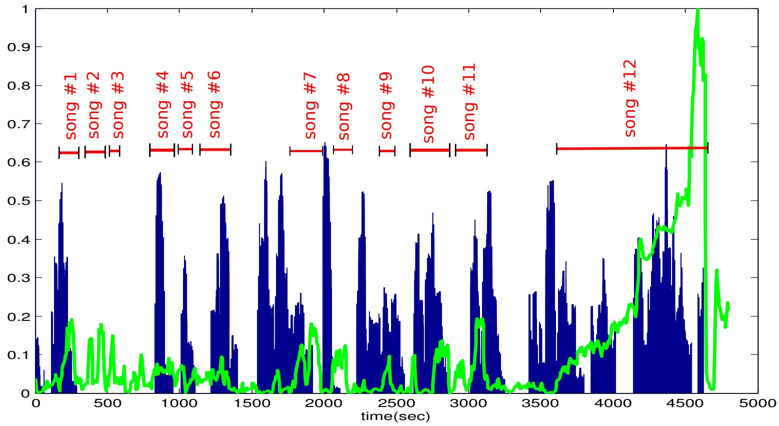


Fig. 7.2 Sound intensity of the performance (green) compared against the normalized co-occurrence measurement calculated by MI (blue).

performance's songs are also highlighted in this figure in red. Here, it can be seen that although the music had a correlation with the response of the public in a performance in certain sequences, other moments of high mean MI are also correlated with acts with no music. This suggests that the music may not have been the main factor stimulating coordination between our participants. For this reason, we decided that the song changes would not be useful for the following classification experiments and have focused solely on acceleration data instead.

With this preliminary qualitative information, the next subsection describes our classification experiments using the four questionnaire tasks mentioned in Section 7.3.1.

7.3.4 Classifying experience

Labeling samples. We use a standard pattern recognition approach to automatically predict the responses of participants. We start by labeling our participants according to the questionnaire answers they gave. For each group of three questions, we obtained one numerical value by averaging and rounding the three answers. This way, we obtain four different labels for our participants, where each label corresponds to the one of the tasks. We divided the participants into two classes for each task corresponding to a "positive" and "negative" report on their experience of the performance. Participants whose averaged answer to any group of questions was below 5 was placed in the negative class for that task, meaning this participant, either did not enjoy the event, would not recommend the performance, did not feel immersed throughout the performance or did not think the performance uplifted their mood. The positive class thus contains participants who gave positive responses to the questions.

In this study, we mainly focus on enjoyment and recommendation, since these two are questions with clearer indications, but still provide results for immersion and the mood changes as we believe they can help in obtaining a general understanding of the performance's effects on the participants. After defining the classes for each task, our class distributions were not always balanced. For "enjoyment" and "recommendation", the majority of participants (26) gave positive answers. 22 participants thought "the performance affected their mood positively". Therefore, for these tasks, the problem becomes challenging for the smaller negative classes. These imbalance in class distributions also affected our choices of performance measure; in addition to accuracy, we also provide the balanced accuracy so each classes contributes equally to the final measure. The distribution for the "immersion" task is more balanced with 17 participants in the positive class.

Methodology. To emphasize the connection between the information contained in the motion data and the participants' experience of the event, in our classification experiments we focus on a simple set of features and a well-understood classifier. More precisely, our features are the variance of acceleration along each axis and of acceleration magnitude, extracted with the aforementioned setup. We selected a Linear SVM as our classifier. Since the number samples is limited, we chose to use a model with few parameters . For evaluating the performance of our method, we used leave-one-out cross validation, training with 31 samples and with the remaining one is used for testing. The hyperparameters of the SVM are also selected using cross validation on the training set.

As stated in Section 7.3.2, representing the features using the whole performance would require classification on a 18820-dimensional feature vector for each participant. Since we do not expect all intervals to be equally informative and to avoid the curse of dimensionality, we decided to use a filtering approach which selects the informative intervals before feature extraction.

To do so, we selected a Dynamic Time Warping (DTW) distance computed over a window (sizes were set from 20 to 60 samples) with a shift size of 1 sample . Similar to the mean MI co-occurrence vector

Method \ BAcc | Acc(%)	Enjoyment	Recommend	Immerse	Mood
DTW IS(20 Sample)	60 | 66*	61 | 78	63 | 63	60 | 63*
DTW IS(40 Sample)	90 | 94**	70 | 81	53 | 53	55 | 66*
DTW IS(60 Sample)	78 | 84**	71 | 84	49 | 50	48 | 63*
Whole Event	36 | 38	74 | 78	51 | 50	38 | 38

(→ p<0.05) (** → p<0.01)*

Tab. 7.1 Performance scores obtained with different methods

used in Section 7.3.3. Our assumption here is that if we select the intervals where the average DTW distance between each pair is significantly higher than the rest, we should end up with time intervals that are more discriminative than the rest. In an ideal scenario, intra-class distances should stay relatively stable throughout the event, so the parts where the average DTW distance between pairs is high correspond to intervals where the intra-class distances are maximized. We hypothesize that using this metric provides better discrimination between classes compared to using mutual information where moments of high mutual information could also correspond to moments where the mean DTW is low and both classes would be almost indistinguishable . Empirical results using a threshold on the mean MI supported this claim, with performance scores significantly lower than the proposed method for the majority of tasks.

Using the same setup explained in Section 7.3.3, we detect the intervals where pairwise DTW distances are significantly higher than the rest and extract variance features for each person from these intervals only. The number of remaining intervals after filtering depends on the window size selection. In our experiments, where we have used windows of 20 to 60 samples, the number of selected intervals ranged from 86 to 830. Finally, after interval selection and feature extraction, we perform further dimensionality reduction by applying principal component analysis (PCA) to the feature vectors. We keep the principal components which preserve the 99 percent variance of features and use them for training and testing our model. This resulted in 12 and 19 resulting feature dimensions.

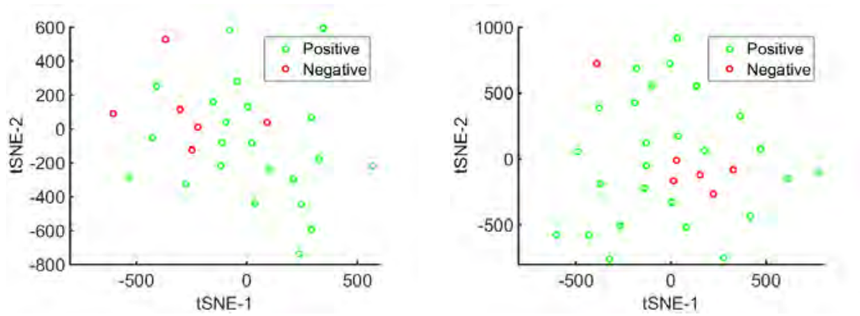
Results. The performance obtained by the proposed method, with different window size selections when using the thresholded DTW distance to pre-filter the salient intervals, are presented in Table 7.1. This table also includes the performance scores obtained without interval selection. The statistically significant results between those using interval selection and those using whole event are specified with a sign '*'.

The performance obtained without interval selection, which are reported in the final row of Table7.1, are unsatisfactory in general. Any task other than predicting recommendation has an accuracy lower than or equal to the proposed method, regardless of the window size. We should note that we did also apply PCA to the feature vectors for the non-filtered method. However, these scores showed that the extracted principal components were still affected by variance features extracted from many non-informative intervals, supporting our claim of interval selection is necessary.

The performance obtained with different window sizes have some interesting implications. For the tasks of immersion and mood, we obtained the highest performance with a window of 20 samples, corresponding to one second. With the increasing window size, our performance for these tasks dropped below random. We can see that this same window size is not optimal for the enjoyment and recommendation tasks. This could suggest that some tasks are shorter in time scale than others. However, we need more data to draw solid conclusions about such implications. It should be also noted that for the mood task, our proposed method always provides a significantly better result than the whole interval method, regardless of the window size.

Another interesting implication can be seen in the performance scores for "recommendation". For this task, the highest performance score is obtained when the whole event was used. Supporting this idea, the performance for this task increases as the window size increases. We should note that using a wider window may not always guarantee that more intervals are selected for classification. None of the results presented for the recommendation task showed significant improvement over the whole event baseline. However, the recommendation accuracy is very high for the whole event, suggesting that the behavior of people who are interested in recommending dance performance to others could be distinguished from those who did not tend to recommend dance performances to others, and that this was independent of any particular moments during a performance.

Finally, we can see that for enjoyment, one of our most important tasks, we obtain relatively high performance, with the highest score of 94% accuracy and 90% balanced accuracy. With a window size of 40 samples, only one sample from each class is misclassified, resulting in high precision and recall scores for both classes. This result is significantly better than the whole interval method ($p>0.01$). To further investigate how the interval selection affects the distribution of sam-



(a) (b)

Fig. 7.3 Non-linear embedding of feature vectors for the enjoyment class for (a) the whole event and (b) using interval selection with a 40 sample window.

ples, we visualize the filtered set of feature vectors of each participant using t-SNE [138], a non-linear embedding technique. The resultant two-dimensional embedding shown in Figure 7.3 illustrates that interval selection allows data points in the negative class to be clustered together in the feature space.

We also experimented with computing DTW distances on the raw accelerometer magnitude signal, instead of the variance. This experiment resulted in performance scores that were worse than random for "enjoyment", "immersion" and "mood". For "recommendation", we obtained a balanced accuracy score of 85 percent. This result is quite interesting, since the highest performance we obtained for "recommendation" in Table 7.1 was the case where the whole event was used. However, in general it can be said that this experiment empirically supported our claim that the variance in acceleration is a valid feature for our experiments, both as a feature and for the interval selection using the thresholded DTW distance.

7.3.5 Further analysis of salient moments

This section aims to further explain the salient moments of the performance, now relating these with the classes identified in Section 7.3.4. For space reasons, we focus on the enjoyment task as it has given us the best performance. The pairwise similarity measurements from the previous qualitative analysis were separated into two groups for each task: the ones who completely agreed with the statement and everyone else. For each group, the same unified similarity measurement as described in Section 7.3.3 was calculated and the salient moments were obtained

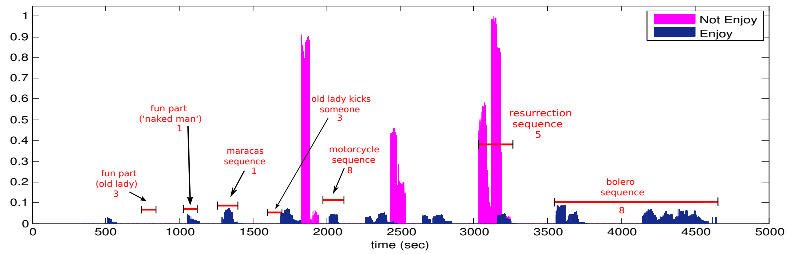


Fig. 7.4 Salient moments from mean MI discriminating people in class 'Enjoy' and 'Not Enjoy'

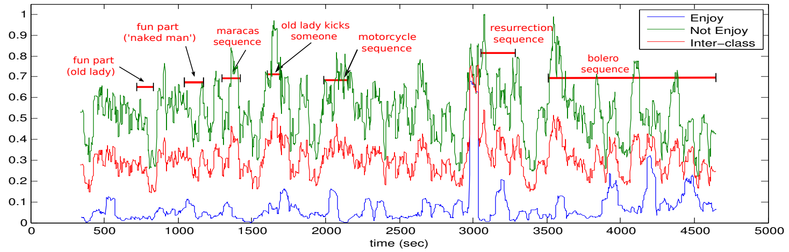


Fig. 7.5 Similarity measurement using DTW for each class in the enjoyment task.

using the Otsu threshold level. Since the goal is to asses the similarity of people within the same class, pairs of different classes are left out.

Figure 7.4 shows the measurements of mean MI over time for both classes in the enjoyment task (where the negative class was plotted below the positive class). Notice how the two moments considered as favorites for the majority of participants (*motorcycle sequence* and *bolero finale*) reappears for the mean MI in the group that enjoyed the performance but not for those who disliked it. Actually, there is almost no overlap between the salient moments for the classes 1 and 2. This reaffirms that specific acts or sequences in a performance (or movie) can have a significant impact in the final assessment of enjoyment.

Furthermore, Figure 7.5 shows the mean DTW distance for members within the 'Enjoy' class (blue), the 'Not Enjoy' class (green) and all pairs in opposing classes (red). The 'Not Enjoy' class resulted in a higher overall DTW distance, over the complete performance, compared to the 'Enjoy' class. This might indicate a lack of synchrony among people who dislike the performance, which echoes findings by Wang and Cesar with Galvanic Skin Response measures to an audience's reaction to a live performance [143].

7.3.6 Identifying sitting neighbors

Knowing who is sitting close to whom is useful to study the spatio-temporal properties of audience response. Do neighboring individuals talk to each other at certain specific moments of the performance? Do people sitting close to each other react the same way (or at the same time)? Does response (e.g., movement, emotion, mood) spread spatially, perhaps due to influence of neighbors? Yet, who is sitting where (or at least close to whom) is often unknown, as the audience is usually free to choose where to seat and only rarely are seats assigned to the tickets by the theater. In this section we investigate whether we can leverage proximity data collected during the performance to identify who is sitting close to whom.

Given that seats are placed such that all the audience is facing towards the same place, the stage, and that the proximity sensor worn on the chest yields face-to-face proximity detections, it is not obvious what kind of information can be extracted from proximity detections collected during the performance. With sensors detecting other sensors at an angle of around 60 degrees towards the front, it is not clear how consistently they can detect sensors worn by individuals sitting at the two sides, and how the shielding effect of the body influences the detections of individuals sitting in front or behind.

Furthermore, even assuming neighbors can be detected, it is unclear how far they can be sensed (e.g., at how many rows or columns of distance), and how this relationship can be characterized, in particular considering no signal-strength (i.e., RSSI) is recorded by the sensors.

To start, one would assume that the closer two individuals sit, within the detection range of the sensor of 2-3 meters, the more frequently their sensors will detect each other. First, given an individual u, we compute the number of times each ID of another sensor is detected during the performance by the sensor worn by u, and we keep the top-K. Then, as we have recorded where volunteers were sitting in the theater in the ground-truth, we generate ground-truth neighbors layout for each individual, and we compute precision and recall of the measurement, for a given K.

Precisely, assuming individual u is sitting at position $u_{i,j}$, meaning at row i and column j, we compute from ground-truth the IDs of the individuals (i) sitting at the sides ($K = 2$ individuals, which we call 1-hop side neighbors, that is at positions $u_{i,j+1}$ and $u_{i,j-1}$) (ii) sitting in front and behind ($K = 2$ individuals, which we call frontal, that is at positions $u_{i+1,j}$ and $u_{i+1,j}$), (iii) sitting at the sides of the 1-hop side neighbors ($K = 2$ individuals, which we call 2-hop neighbors, that is at positions $u_{i,j-2}$ and $u_{i,j+2}$), and (iv) sitting at "diagonals" ($K = 4$ individuals, that is at positions $u_{i-1,j-1}$, $u_{i-1,j+1}$, $u_{i+1,j+1}$, and $u_{i+1,j-1}$). We then compare these lists of IDs with those in the top-K lists.

Frontal and diagonal neighbors yield low recalls of 0.37 and 0.24 respectively, while 1-hop neighbors yield precision of 0.62 and recall of 0.86. When we add also 2-hop neighbors, and hence using a value of $K = 4$ (that is, we consider the 4 IDs appearing most frequently to detect 1-hop and 2-hop neighbors, 4 individuals in total), we obtain a precision of 0.59 and a recall of 0.84. These results suggest that some of the neighbors detected for the 1-hop neighbors (with $K = 2$) are 2-hop neighbors (lowering the precision), but 2-hop neighbors are not consistently detected such that precision and recall are still similar with $K = 4$. The other source of error in precision in both cases are the rare detections of frontal and diagonal neighbors, which are not detected consistently but sometimes appear in the top-K list for some individuals. Moreover, the fact that 1-hop and 2-hop neighbors are mixed in the top-K list makes it impossible to identify which is which based on frequency (i.e., if 1-hop neighbors were detected more frequently than 2-hop neighbors it would trivial).

Finally, one should notice that the K parameter is applied to all individuals in the same way. However, for individuals who sit at the sides of the theater, and hence have fewer 1-hop and 2-hop neighbors, the fixed value is not optimal. We investigated whether we could detect these side cases by looking at the frequency distribution, but again we found that there was no consistent way of classifying in which position a neighbor was sitting, given that neighbors are detected with a frequency not directly connected to their position.

To conclude, the precision and recall values obtained when classifying 1-hop and 2-hop neighbors show that it is possible to detect who is sitting at the sides of an individual (with some sampling of frontal and diagonal neighbors), and we plan to investigate further how signal-strength can help disambiguate between these classes.

7.4 Case study 2: impact of a performance on social behavior

Section 7.3 provided interesting insights into how the response to a dance performance can be measured with pervasive sensing. However, while working with HD, we came across a different perspective on

Fig. 7.6 Snapshots of the instrumented mingling room.

the problem. Can we quantify the influence of a performance on the audience even after the performance is over?

There is clearly a context that surrounds the event itself — typically, people will attend a performance with friends and/or family, may come for a drink beforehand and stay for drink afterwards. We hypothesized that people's social behavior (as measured through proximity and acceleration) could also be affected by watching a dance performance. As a business model HD was already co-organizing networking events around dance performances together with two local networking organizations. The idea was that the dance performance could be an occasion to enhance the networking event, and the co-located networking event would encourage more people to watch dance.

To investigate this hypothesis, we decided to investigate whether we could measure differences in how people socialized during the event. Hence, we measured mingling behavior during two networking sessions, one right before a dance performance and another right after it. With HD and regional networking groups, we co-organised a networking event with 48 volunteers. The same sensing devices described in Section 7.3 were used. An example snapshot of the mingling data is shown in Figure 7.6.

Although a networking event is not exactly the same as the more casual ways that people might attend dance performances socially, we believe this initial investigation provides a feasibility study for larger scale less controlled studies in the future.

Similar to previous work using proximity sensors to analyze social behavior in conferences [97] (see Chapter 3), musea [26], and workplaces [110], we used proximity sensors as proxies for face-to-face social interactions, together with accelerometer data as described above. Two months after the experiment, we published sensor analysis results for the volunteers and asked them to answer a survey about their experience of the dance performance and the networking event.

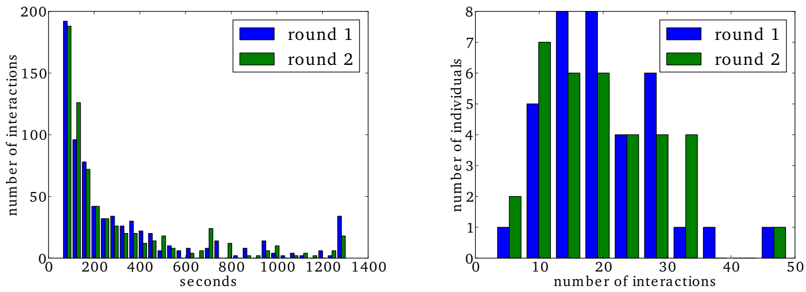
7.4.1 Setup

To measure whether people were interacting or not, we processed the proximity detections collected by the devices as follows. Because of the unreliability of the wireless medium, we used a density-based filtering technique to increase the sensitivity of the signal for detecting face-to-face proximity [98] (see Chapter 4). For each pair of individuals, we computed the intervals where pairs were continuously facing each other, formally $[t_i, t_j]$ where t_i is the timestamp of the first detection and t_j is the timestamp of the last detection of the interval. Because pairs can be close for multiple non-overlapping time intervals during the same measurement, we computed multiple intervals for the same pair. Here, we refer to an interval of proximity between any pair as an *interaction*. For our experiments we considered only intervals of proximity longer than 60s to indicate interactions.

7.4.2 Results

Since some attendees did not attend the entire event, only the data from 35 of the participants was available for analysis. We first hypothesized a difference in the length or the number of interactions between the two sessions. For example, one could imagine that individuals would interact in longer conversations, or with more people. In Figure 7.7(a) we present the distribution of the length of the interactions for the two sessions (from here on referred to as round 1 and round 2) across all the individuals. In both rounds shorter interactions are predominant. Note that as drinks were served at the bar, during both rounds often individuals left a conversation to fill their glass and went back right afterwards to the same conversation, which would be measured as two distinct interactions. No significant mean difference was seen between the distribution in interaction length for the two rounds. In Figure 7.7(b) we present the distribution of the number of distinct interactions for round 1 and round 2 across all the individuals.

A second difference we hypothesized was in the size of conversational groups. For example, people could be engaged in conversations involving more people, or conversely more one-to-one conversations, perhaps to discuss the content of the performance. We define a *neighborhood* as the set of nodes a sensor *a* detects at a given moment in time, i.e.,



(a) interaction lengths

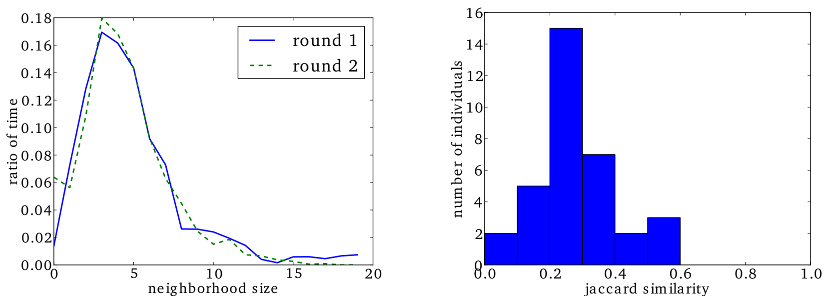
(b) interaction numbers

Fig. 7.7 (a) Distribution of the lengths of the interactions during the two rounds. (b) Distribution of the number of interactions for round 1 and round 2 across all the individuals.

the individuals in physical proximity of the individual wearing sensor a. In Figure 7.8(a) we present the distribution of neighborhood size with respect to the amount of time they were observed together, expressed as a ratio over the round duration. In other words, it represents the amount of time individuals have spent in proximity to another n individuals. The results show a peak around four individuals, a reasonable group size for a conversation. Similar to the interaction lengths, the two distributions look very similar.

The third hypothesis regarded changes in conversational partners. For example, people could be interacting with the same individuals as before the performance, or be stimulated to engage with others. To this end, for each individual we computed the *Jaccard similarity* between the set of participants an individual has interacted with during the two rounds. Given two sets of IDs R_1 and R_2, the Jaccard similarity function is defined as $J(R_1,R_2) = \frac{|R_1 \cap R_2|}{|R_1 \cup R_2|}$ and computes a value in the interval $[0,1]$. Figure 7.8(b) presents the distribution of the Jaccard similarity across all individuals between round 1 and round 2. The results show that although the mingling pattern of the individuals did not change between the two rounds, they did interact with different individuals. In particular, they changed at least 50% of their interaction partners between round 1 and round 2 (mean 0.278 and standard deviation 0.121).

Acceleration. The image emerging from the pure proximity measurement is that of an ordinary mingling event. Overall, these results indicate



(a) conversational group sizes (b) change in partners

Fig. 7.8 (a) Relative amount of time sensors detected a certain number of other sensors (at a specific moment in time). (b) Distribution of the Jaccard similarity across the individuals.

that the volunteers, as a group, applied a consistent pattern in their mingling behavior during the two rounds, a pattern that they used, however, to target different conversational partners between the two rounds. The measurement pictures a socializing context, but it is difficult to reach conclusions about the impact of the performance. For this reason, we focused on the acceleration data as well.

Similar to the direct approach, we used variance in the acceleration magnitude as the main feature. We then correlated the participant's self-reported behavior with our findings. Correlation between the answers to question "Do you think the performance had an effect on your mood? Yes|No" and the difference between the acceleration magnitude variance in round 1 and 2 is computed. Since not all participants filled in the post event survey and some accelerometers failed due to a firmware bug, we were only able to use the accelerometer data from 14 participants.

The variance values are extracted using the whole intervals for round 1 and 2. A statistically significant ($p = 0.02$) positive correlation value of 0.60 was obtained as a result. This correlation supports our hypothesis that the mood change can be linked to implicit behavior as measured by acceleration, though we would like to verify this with a larger dataset later. In conclusion, the results suggest that while individuals acted similarly as a group in terms of networking behavior captured by the proximity sensors, the quality of those interactions seemed different between the two sessions, as captured by the accelerometers.

7.5 Discussion

We have shown that by leveraging variance in the accelerometric signal it is possible to predict the response of the audience of a live dance performance. One of the main contributions of such result is that it shows that information about the internal experience of the audience is embedded into something as simple as the small movements of seated individuals, if the whole involved crowd is taken into account. While very promising, much more work needs to be done.

First, we have employed a blackbox supervised approach, where we predicted the labels collected in the surveys. While labels could be collected for a subset of the audience and a semi-supervised approach could be followed, it is not realistic to believe that one may require to collect labels for each performance. Unsupervised approaches should be investigated further, and most importantly it is necessary to understand what fundamental and principal elements of the signal relate to which moods, emotions, and in general qualities of the response, and in which way.

Second, we have focused on summarizing the whole experience with a label, whereas such method could, at least in principle, focus on specific moments in time, hence applying different labels at different times. Being able to characterize certain emotions and responses during specific moments of an event, and as early as possible, should enable managers to predict and act in a timely fashion, to avoid accidents and steer a crowd towards safety and comfort.

Third, our volunteers were sitting during the performance, so more research needs to be done to understand how to apply the methodology to different crowds and conditions where individuals are standing, moving or even dancing. The seated state has made things both simpler and harder at the same time. On the one hand, the range of ancillary behaviors and actions have been minimized to the minimum, with the individuals concentrating into following the dance performance at all times. On other hand, the range of movements have also been reduced to small and short movements, removing possibly more explicit reactions that are possible when an individual is free to move in space, and that would emerge, for example, in cases of emergency. Different events, conditions, and responses need to be investigated in the future.

Fourth, with the upcoming wave of wearables and sensors, accelerometers are just an example of sensors that can be used to collect data about crowd response in a pervasive way. For example, new generations of smart watches already employ galvanic skin response sensors, which should yield information about arousal over time. We need to investigate which signal and sensor is more appropriate in which condition, and whether fusing the sources of data can improve the picture we obtain out of the measurement.

Finally, regarding proximity, we have shown how spatio-temporal information can be used to understand who is seated close to whom and who is talking to whom. However, we have only started to scratch the surface of how spatio-temporal information, and in particular the temporal dimension, can be used to characterize the nature of the response of the crowd, how it spreads, and how the different environmental and internal factors influence the response.

7.6 Conclusions

We have made a first investigation of how a seated audience's perception of a performance can be perceived and measured from their body movements using an accelerometer that exists typically in smart phones. We analyzed whether subtle and complex concepts such as "enjoyment", "immersion", an in improvement in mood as results of the performance, and whether participants would "recommend" dance in general, would be reflected in body motion using a simple accelerometer hung around the neck. Using the variance of the acceleration, we were already able to predict with 90% accuracy whether somebody enjoys the performance, which performed significantly above a random baseline.

Importantly, joint coordination in the variance in acceleration helps to distinguish salient from non-salient moments, which lead to significant improvements over using each person's body movements from the entire performance period. As well as the obvious usefulness to the entertainment industry of such direct measurements of an audience's reaction, we have also made a first attempt to measure the role that a live performance can have on the social behavior that precedes and follows it. Our experiments shows promise in enabling us to measure the implicit responses of people while watching a live performance without the need for more traditional sensing approaches using physiological or brain signals.

8 Conclusions

> *Tell me and I forget, teach me and I may remember, involve me and I learn.*
>
> — **Benjamin Franklin**

Spatio-temporal proximity data about individuals and objects can give insights about our social behavior and open a window in the digital world over our life in the real world. Crowd managers report lack of situation awareness in their processes, and a desire for stronger ability to monitor and predict the behavior of a crowd, as well as to intervene to steer the crowd to guarantee safety and comfort. Inexpensive miniaturized proximity sensors, both mobile and fixed, can be used to collect such spatio-temporal proximity information continuously, as the basis for tools to fulfill such needs.

Due to the unreliability of wireless communication, the data provided by the sensors needs to be filtered. Once the data is collected and filtered, spatio-temporal proximity data can be modeled as a series of so-called proximity graphs. By analyzing series of proximity graphs it is possible to identify and characterize different crowd behaviors, and better understand what is happening in a crowd. It is important for these tools to operate in extreme cases, for example, of high density and challenging environmental conditions, that is when such tools are most needed. Moreover, spatio-temporal proximity information can be used in cooperation with other sensor data to unveil higher-level aspects of crowd behavior, including the more internal experience of the individuals.

In this dissertation we have presented approaches to tackle all these tasks, and that is how spatio-temporal proximity information can be collected, filtered, analyzed, and visualized to gain a better understanding of the behavior of a crowd.

8.1 Summary of contributions

In particular, this dissertation makes the following contributions.

In Chapter 2, we studied current crowd management practices by means of ten interviews conducted with professionals managing some of the largest crowds in The Netherlands. Needs and requirements were identified, including the limitations of current approaches and technologies. In particular, the interviewees reported a strong need for increased situation awareness, prediction, and intervention, positioning spatio-temporal aspects of crowd behavior, like density, flows, and congestions, at the top of their priorities. Recommendations were provided within the framework of a techno-social system for crowd management comprising two feedback-control loops.

In Chapter 3, we dived into the definitions and properties of crowd textures and proximity graphs. Examples of analyses to identify pedestrian lanes, bottlenecks, and social groups were provided, along with a discussion about the importance of taking into account the context around a crowd, as well as advantages of instrumenting objects as well as individuals with proximity sensors. A real-world use case was presented, by identifying social groups within face-to-face proximity from data collected during an ICT conference.

In Chapter 4, we analyzed some of the sources of unreliability in mobile and fixed proximity sensing when applied to crowd monitoring. A filter with a few parameters, and a technique to choose good values for them, was presented and evaluated both in simulation and in a number of real-world experiments. The technique can double the sensitivity of the measurement, reconstructing nearly all missed proximity detections and ignoring spurious ones.

In Chapter 5, we focused on tracking an indoor crowd inside a museum, studying in particular person-to-object face-to-face proximity relationships. A pipeline of filters was presented, comprising a particle filter adapted to directional sensing, together with a number of applications to visualize, cluster, and predict the behavior of the museum visitors, with data collected during a real-world 5-days experiment conducted at the CoBrA museum of Amsterdam. We also looked into the results of a number of interviews and a focus group conducted with museum staff regarding the effectiveness of the visualizations.

In Chapter 6, we looked into an extension of the previous technique that takes into account also person-to-person proximity, to overcome limitations caused by high densities. The technique was evaluated during a large-scale experiment with nearly a thousand participants conducted at the NEMO museum, a particularly challenging setup due to the peculiar features of the NEMO building. Together with the filtering pipeline, an analysis to detect crowded events as well as flows between floors was presented.

In Chapter 7, we studied how proximity sensors can be used together with accelerometers to study the response of a crowd to a dance performance. An experiment was presented where audience behavior was tracked before, during, and after the performance, to try to identify the response of the audience (e.g., enjoyment) by looking into the collective movements of the (seated) audience during the performance, and the differences in mingling behavior exhibited before and after the performance.

8.2 Limitations and future research

To conclude, we look at a number of possible directions for future research, bootstrapped by some strengths and limitations of the work presented in this dissertation.

Fine-grained crowd dynamics. In Chapter 2 we have seen how crowd managers put increasing situation awareness about crowd dynamics and all behaviors with strong spatio-temporal nature at the top of their priorities. In Chapter 6 we have focused on detecting density spots and flows, but a more fine-grained approach, for example, like those described in Chapter 3 to detect congestions and pedestrian lanes, should provide more detailed information about the properties of these crowd dynamics. While quantifying the number of people moving between floors at each moment helped unveiling the impact of the Chain Reaction event at the NEMO museum, understanding whether lanes form in these flows in an orderly fashion, as opposed to being perturbed and impeded by environmental factors, would enable crowd managers to react and influence the design of the museum (e.g., the stairs, the positions of some exhibits and divided walls, etc.). The information in the series of proximity graphs should allow deeper analyses of the spatio-temporal relationships between the crowd members to identify and characterize even more precisely a wide range of existing crowd dynamics, and perhaps be the basis for the discovery of new ones.

Integration into wearables and IoT. In this work, we have utilized a number of custom-made devices that were either worn on the chest hanging from the neck or as bracelet. For the former, we have exploited the shielding effect of the body for face-to-face proximity detection, while the latter provided a more uniform area of sensing around the individual. We have also assumed that similar sensing capabilities could be obtained with smart phones via BLE, perhaps placed in a jacket pocket, or with a smart watch. These assumptions are not met, for example, when users put the phone in the trouser's front or back pocket, or worse in a bag. It is important to find a consistent, reliable, and controllable way to shape the area of detection of the proximity sensor around the individual. A promising new area of research is the integration of sensor and computing technology into clothing and textiles, where "wires" are intermixed with standard clothing materials in an imperceivable way, opening to a number of possible application for both sensing, body-area networks, and interactive interfaces. Integrating proximity sensors, for example, inside of cloths like a jacket would allow to obtain a higher level of control about where the sensor is positioned with respect to the body of the individual. A similar argument can be made regarding the integration into objects and things. We interact in a different way with different objects like, for example, a fridge or a book, and different approaches need to be found to be able to capture the nuances of each specific interaction type.

Sensing technology. BLE-like and in general radio-based proximity sensors are just examples of ways to detect proximity between two devices. Each specific technology comes with a different set of advantages and disadvantages that can be exploited to obtain particular features for the proximity sensor. For example, radio-based proximity sensing is known to be influenced by the water and human bodies, but can pass through certain types of materials that are used, for example, to build walls. This can make sensing co-location in dense environments particularly challenging. On the other hand, light-based sensing tends to be easier to confine within the boundaries of rooms and spaces. Furthermore, even remaining within radio-based proximity sensing, the choice of frequency can have a large impact. For example, frequencies around the 2.4Ghz band such as those used by BLE and Zigbee tend to be particularly affected by water, whereas lower frequencies around the 400Khz band, like those used by avalanche transceivers, are less affected by water but afford lower bandwidth. Further research is required into mixing different proximity sensing technologies and signals at the same time, to exploit the respective advantages and overcome the disadvantages (i.e., via multi-medium sensing). Finally, if the approaches presented in this dissertation were really to be deployed large-scale into consumer products like smart phones, more energy-efficient techniques would be necessary, as scanning BLE devices every second would drain the battery of a smart phone within minutes.

Local and hybrid localization. Clearly, there are many points of overlap between proximity-based positioning systems and indoor localizations systems. One could consider in fact the system presented in Chapter 6 as a *local* localization system, where anchors act as landmarks for specific local areas of interest (though we put less impact on computing coordinates). The advantage is the ability to control accuracy in certain areas that are more challenging, for example, due to density or environmental conditions, by adding more anchors were needed. Upcoming extensions to BLE including angle-of-arrival information should allow beacons to be utilized in the future as low-cost solutions for landmarking as in traditional indoor localization systems where needed, or as positioning anchors as proposed in this dissertation, hence forming hybrid solutions tailored to the specific needs of the user.

Feedback and steering of behavior. In this dissertation we have positioned our work within a framework based on feedback-control loops. Large parts of our contributions focus on the sensing/input and processing/control steps of the loop, while we have implemented the feedback step through visualizations. At this time, there is little knowledge about this latter step of the loop and research effort should be directed towards research questions regarding, for example, (i) what information is required by decisions makers at all levels (hence, including both managers and crowd members) and how it should be presented to them, (ii) what intervention strategies are effective in which situations, (iii) how to predict and avoid conditions instead of fixing them, as pro-active avoidance is favorable to reaction, (iv) how technology can help to implement intervention strategies, for example, through actuators, interaction interfaces, data sharing platforms etc. This work requires a multi-disciplinary approach involving the social and behavioral sciences, human-computer and user experience expertise, physics and mathematical modeling, and information-technology engineering. This is perhaps, at this point, the most challenging step in future work.

Applications beyond crowd behavior. In this work, we have focused on crowd behavior as an instance of social behavior. However, with beacons deployed and detectable at large-scale in cities and buildings, the smart devices we wear every day can act as HTTP/Web cookies of the physical world. The number of applications that constant tracking of our whereabouts enables are countless, with even more effective advertising, to continue the analogy with HTTP/Web cookies, as the most clear example. Tracking proximity between individuals and objects is at the basis of paradigms like the IoT and in general location- and context-aware computing, and one can easily imagine impact on transport and logistics, product placement and recommendation in retail, support of social and entertainment events, payments. Clearly, a wide adoption of such monitoring and tracking applications requires devising techniques to maintain and respect the privacy of the users at all times, another important direction for future work.

Wider multi-sensor approach. As discussed in this dissertation, spatio-temporal proximity can only go so far to support the understanding of the interactions between individuals (and objects). The realm of the study of the quality of these interactions belongs to other type of sensors, operating as gateways to the internal experience of the individuals. Smart watches, for example, are now provided with galvanic skin response sensors, heartbeat and blood pressure monitors, microphones and cameras, in addition to the sensors used to track activities, like accelerometers and gyroscopes. The value of the data provided by these sensors goes beyond the monitoring of health levels. As shown in Chapter 7, this data contains information about the response of the individuals, and proximity information can act as the foundations to collect information about the context in which the individual is embedded during those responses tracked by the other sensors. Future research is required to identify which sensor is most effective in which context, perhaps also in conjunction with activity monitoring sensors that further enhance the context recognition capabilities of our devices. Finally, the information they provide can be fused together, including that of other individuals, to finally provide more complete and aware support.

Bibliography

[1] Nedal Ababneh. „Radio irregularity problem in wireless sensor networks: New experimental results“. In: *Sarnoff Symposium, 2009. SARNOFF'09. IEEE*. IEEE. 2009, pp. 1–5 (cit. on pp. 70, 92).

[2] JL Abbott and Morgan W Geddie. „Event and venue management: Minimizing liability through effective crowd management techniques“. In: *Event Management* 6.4 (2000), pp. 259–270 (cit. on p. 12).

[3] Norman Abramson. „The throughput of packet broadcasting channels“. In: *IEEE Transactions on Communications* 25.1 (1977), pp. 117–128 (cit. on p. 69).

[4] S Adler, S Schmitt, Y Yang, Y Zhao, and M Kyas. *Fubloc: Accurate range-based indoor localization and tracking*. Tech. rep. Technical report, Microsoft Indoor Localization Competition, 2014 (cit. on p. 90).

[5] Rakesh Agrawal and Ramakrishnan Srikant. „Privacy-preserving data mining“. In: *ACM Sigmod Record*. Vol. 29. 2. ACM. 2000, pp. 439–450 (cit. on p. 37).

[6] Nadav Aharony, Wei Pan, Cory Ip, Inas Khayal, and Alex Pentland. „Social fMRI: Investigating and shaping social mechanisms in the real world“. In: *Pervasive and Mobile Computing* 7.6 (2011), pp. 643–659 (cit. on p. 59).

[7] S Al Bosta. „Crowd management based on scientific research to prevent crowd panic and disasters“. In: *Pedestrian and Evacuation Dynamics*. Springer, 2011, pp. 741–746 (cit. on p. 13).

[8] Nils Aschenbruck, Raphael Ernst, Elmar Gerhards-Padilla, and Matthias Schwamborn. „BonnMotion: a mobility scenario generation and analysis tool“. In: *Proceedings of the 3rd International ICST Conference on Simulation Tools and Techniques*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering). 2010, p. 51 (cit. on p. 68).

[9] Ashwin Ashok, Chenren Xu, Tam Vu, et al. „Bifocus: Using radio-optical beacons for an augmented reality search application“. In: *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*. ACM. 2013, pp. 507–508 (cit. on p. 90).

[10] Eleana Asimakopoulou and Nik Bessis. „Buildings and crowds: Forming smart cities for more effective disaster management“. In: *Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2011 Fifth International Conference on*. IEEE. 2011, pp. 229–234 (cit. on p. 14).

[11] Louis Atallah and Guang-Zhong Yang. „The use of pervasive sensing for behaviour profiling - a survey“. In: *Pervasive and Mobile Computing* 5.5 (2009), pp. 447–464 (cit. on p. 13).

[12] Ling Bao and Stephen S Intille. „Activity recognition from user-annotated acceleration data“. In: *International Conference on Pervasive Computing*. Springer. 2004, pp. 1–17 (cit. on p. 146).

[13] Xuan Bao, Songchun Fan, Alexander Varshavsky, Kevin Li, and Romit Roy Choudhury. „Your reactions suggest you liked the movie: Automatic content rating via reaction sensing“. In: *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. ACM. 2013, pp. 197–206 (cit. on p. 147).

[14] Christian Beder and Martin Klepal. „Fingerprinting based localisation revisited: A rigorous approach for comparing RSSI measurements coping with missed access points and differing antenna attenuations“. In: *Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on*. IEEE. 2012, pp. 1–7 (cit. on pp. 90, 125).

[15] Nicola Bellomo and Christian Dogbe. „On the modeling of traffic and crowds: A survey of models, speculations, and perspectives“. In: *SIAM review* 53.3 (2011), pp. 409–463 (cit. on pp. 13, 34).

[16] Lucy Bennett. „Patterns of listening through social media: online fan engagement with the live music experience“. In: *Social Semiotics* 22.5 (2012), pp. 545–557 (cit. on p. 145).

[17] Richard Bentley, John A Hughes, David Randall, et al. „Ethnographically-informed systems design for air traffic control“. In: *Proceedings of the 1992 ACM conference on Computer-supported cooperative work*. ACM. 1992, pp. 123–129 (cit. on p. 14).

[18] Alexander E Berlonghi. „Understanding and planning for different spectator crowds“. In: *Safety Science* 18.4 (1995), pp. 239–247 (cit. on pp. 19, 38).

[19] Ulf Blanke, Gerhard Troster, Tobias Franke, and Paul 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*. IEEE. 2014, pp. 1–7 (cit. on p. 35).

[20] Bettina Bläsing, Beatriz Calvo-Merino, Emily S Cross, et al. „Neurocognitive control in dance perception and performance“. In: *Acta psychologica* 139.2 (2012), pp. 300–308 (cit. on p. 145).

[21] Alan S Brown and Jennifer L Novak. *Assessing the intrinsic impacts of a live performance*. WolfBrown San Francisco, CA, 2007 (cit. on p. 145).

[22] Erich Bruns, Benjamin Brombach, Thomas Zeidler, and Oliver Bimber. „Enabling mobile phones to support large-scale museum guidance“. In: *IEEE multimedia* 14.2 (2007), pp. 16–25 (cit. on pp. 91, 125).

[23] Tung X Bui and Siva R Sankaran. „Design considerations for a virtual information center for humanitarian assistance/disaster relief using workflow modeling“. In: *Decision support systems* 31.2 (2001), pp. 165–179 (cit. on p. 14).

[24] Beatriz Calvo-Merino, Corinne Jola, Daniel E Glaser, and Patrick Haggard. „Towards a sensorimotor aesthetics of performing art“. In: *Consciousness and cognition* 17.3 (2008), pp. 911–922 (cit. on p. 145).

[25] Marco Cattani, Marco Zuniga, Andreas Loukas, and Koen Langendoen. „Lightweight neighborhood cardinality estimation in dynamic wireless networks“. In: *Proceedings of the 13th international symposium on Information processing in sensor networks*. IEEE Press. 2014, pp. 179–189 (cit. on p. 33).

[26] Ciro Cattuto, Wouter Van den Broeck, Alain Barrat, et al. „Dynamics of person-to-person interactions from distributed RFID sensor networks“. In: *PloS one* 5.7 (2010), e11596 (cit. on p. 159).

[27] R Challenger, CW Clegg, and MA Robinson. „Understanding crowd behaviours, Tech. rep“. In: *Understanding Crowd Behaviours (Crown, 2009)* (2009), pp. 1–326 (cit. on p. 34).

[28] Rose Challenger and Chris W Clegg. „Crowd disasters: a socio-technical systems perspective“. In: *Contemporary Social Science* 6.3 (2011), pp. 343–360 (cit. on p. 12).

[29] W Challenger, WC Clegg, and AM Robinson. „Understanding crowd behaviours: Guidance and lessons identified“. In: *UK Cabinet Office* (2009) (cit. on pp. 11, 13, 38).

[30] Albert Cherns. „Principles of sociotechnical design revisted“. In: *Human relations* 40.3 (1987), pp. 153–161 (cit. on p. 14).

[31] Albert Cherns. „The Principles of Sociotechnical Design1“. In: *Human relations* 29.8 (1976), pp. 783–792 (cit. on p. 14).

[32] Tanzeem Choudhury and Alex Pentland. „Sensing and modeling human networks using the sociometer“. In: *interactions* 5 (2003), p. 6 (cit. on pp. 42, 48, 80).

[33] Cinimodstudio. *Peru national football stadium*. `http://cinimodstudio.com/project/peru-national-football-stadium/`. 2011 (cit. on p. 13).

[34] Chris Clegg and Craig Shepherd. „'The biggest computer programme in the world ever!': time for a change in mindset?“ In: *Journal of Information Technology* 22.3 (2007), pp. 212–221 (cit. on p. 14).

[35] Chris W Clegg. „Sociotechnical principles for system design“. In: *Applied ergonomics* 31.5 (2000), pp. 463–477 (cit. on p. 14).

[36] Claudine Conrado and Patrick de Oude. „Scenario-based reasoning and probabilistic models for decision support“. In: *Information Fusion (FUSION), 2014 17th International Conference on*. IEEE. 2014, pp. 1–9 (cit. on p. 34).

[37] Lynne Cooke. „Assessing concurrent think-aloud protocol as a usability test method: A technical communication approach“. In: *Professional Communication, IEEE Transactions on* 53.3 (2010) (cit. on p. 108).

[38] Emily S Cross, Louise Kirsch, Luca F Ticini, and Simone Schütz-Bosbach. „The impact of aesthetic evaluation and physical ability on dance perception“. In: *Frontiers in human neuroscience* 5 (2011) (cit. on p. 145).

[39] Fred D Davis, Richard P Bagozzi, and Paul R Warshaw. „User acceptance of computer technology: a comparison of two theoretical models“. In: *Management science* 35.8 (1989), pp. 982–1003 (cit. on p. 36).

[40] Matthew C Davis, Rose Challenger, Dharshana NW Jayewardene, and Chris W Clegg. „Advancing socio-technical systems thinking: A call for bravery“. In: *Applied ergonomics* 45.2 (2014), pp. 171–180 (cit. on p. 14).

[41] JF Dickie. „Major crowd catastrophes“. In: *Safety science* 18.4 (1995), pp. 309–320 (cit. on p. 12).

[42] Eyal Dim and Tsvi Kuflik. „Automatic detection of social behavior of museum visitor pairs“. In: *ACM Transactions on Interactive Intelligent Systems (TiiS)* 4.4 (2015), p. 17 (cit. on pp. 91, 125).

[43] Trinh Minh Tri Do and Daniel Gatica-Perez. „Human interaction discovery in smartphone proximity networks“. In: *Personal and Ubiquitous Computing* 17.3 (2013), pp. 413–431 (cit. on p. 59).

[44] Matthew Dobson, Spyros Voulgaris, and Maarten van Steen. „Merging ultra-low duty cycle networks“. In: *2011 IEEE/IFIP 41st International Conference on Dependable Systems & Networks (DSN)*. IEEE. 2011, pp. 538–549 (cit. on pp. 53, 60, 68).

[45] Matthew Dobson, Spyros Voulgaris, and Maarten Van Steen. „Network-level synchronization in decentralized social ad-hoc networks“. In: *Pervasive Computing and Applications (ICPCA), 2010 5th International Conference on*. IEEE. 2010, pp. 206–212 (cit. on pp. 60, 68).

[46] Charalampos Doukas, Ilias Maglogiannis, Philippos Tragas, Dimitris Liapis, and Gregory Yovanof. „Patient fall detection using support vector machines“. In: *IFIP International Conference on Artificial Intelligence Applications and Innovations*. Springer. 2007, pp. 147–156 (cit. on p. 146).

[47] J Drechsel and A Kimms. „Task force deployment for big events“. In: *Safety science* 46.9 (2008), pp. 1289–1305 (cit. on p. 36).

[48] Dorine C Duives, Winnie Daamen, and Serge P Hoogendoorn. „State-of-the-art crowd motion simulation models“. In: *Transportation research part C: emerging technologies* 37 (2013), pp. 193–209 (cit. on p. 13).

[49] John RG Dyer, Anders Johansson, Dirk Helbing, Iain D Couzin, and Jens Krause. „Leadership, consensus decision making and collective behaviour in humans“. In: *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 364.1518 (2009), pp. 781–789 (cit. on pp. 42, 45).

[50] Nathan Eagle and Alex Sandy Pentland. „Reality mining: sensing complex social systems“. In: *Personal and ubiquitous computing* 10.4 (2006), pp. 255–268 (cit. on p. 59).

[51] Cameron Earl, Elizabeth Parker, Andrew Tatrai, Mike Capra, et al. „Influences on crowd behaviour at outdoor music festivals“. In: *Environmental Health* 4.2 (2004), p. 55 (cit. on p. 12).

[52] Mica R Endsley. „Toward a theory of situation awareness in dynamic systems“. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37.1 (1995), pp. 32–64 (cit. on p. 29).

[53] Gwenn Englebienne and Hayley Hung. „Mining for motivation: using a single wearable accelerometer to detect people's interests“. In: *Proceedings of the 2nd ACM international workshop on Interactive multimedia on mobile and portable devices*. ACM. 2012, pp. 23–26 (cit. on p. 146).

[54] Zeki Erkin, Jie Li, Arnold POS Vermeeren, and Huib de Ridder. „Privacy-Preserving Emotion Detection for Crowd Management“. In: *Active Media Technology*. Springer, 2014, pp. 359–370 (cit. on p. 37).

[55] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. „A density-based algorithm for discovering clusters in large spatial databases with noise." In: *Kdd*. Vol. 96. 34. 1996, pp. 226–231 (cit. on pp. 64, 95).

[56] Frédéric Evennou, François Marx, and Emil Novakov. „Map-aided indoor mobile positioning system using particle filter". In: *IEEE Wireless Communications and Networking Conference, 2005*. Vol. 4. IEEE. 2005, pp. 2490–2494 (cit. on pp. 93, 128).

[57] Z Fang, SM Lo, and JA Lu. „On the relationship between crowd density and movement velocity". In: *Fire Safety Journal* 38.3 (2003), pp. 271–283 (cit. on p. 29).

[58] Julien Fleureau, Philippe Guillotel, and Izabela Orlac. „Affective benchmarking of movies based on the physiological responses of a real audience". In: *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*. IEEE. 2013, pp. 73–78 (cit. on p. 146).

[59] Behaviour in FTRE. „Human behaviour in fire: The development and maturity of a scholarly study area". In: *Fire Mater* 23 (1999), pp. 249–253 (cit. on p. 35).

[60] Raghu K Ganti, Fan Ye, and Hui Lei. „Mobile crowdsensing: current state and future challenges". In: *Communications Magazine, IEEE* 49.11 (2011), pp. 32–39 (cit. on p. 14).

[61] Raghu K Ganti, Nam Pham, Yu-En Tsai, and Tarek F Abdelzaher. „PoolView: stream privacy for grassroots participatory sensing". In: *Proceedings of the 6th ACM conference on Embedded network sensor systems*. ACM. 2008, pp. 281–294 (cit. on p. 37).

[62] Rafael C Gonzalez and Richard E Woods. „Digital image processing". In: *Nueva Jersey* (2008) (cit. on p. 150).

[63] Fredrik Gustafsson, Fredrik Gunnarsson, Niclas Bergman, et al. „Particle filters for positioning, navigation, and tracking". In: *IEEE Transactions on signal processing* 50.2 (2002), pp. 425–437 (cit. on pp. 93, 128).

[64] S Gwynne, ER Galea, M Owen, Peter J Lawrence, L Filippidis, et al. „A review of the methodologies used in evacuation modelling". In: *Fire and Materials* 23.6 (1999), pp. 383–388 (cit. on pp. 34, 35).

[65] Health and Safety Executive. „Managing Crowds Safely, A guide for organisers at events and venues". In: HSE books, 2014 (cit. on pp. 13, 35, 38).

[66] Dirk Helbing and Anders Johansson. *Pedestrian, crowd and evacuation dynamics*. Springer, 2011 (cit. on p. 43).

[67] Dirk Helbing and Peter Molnar. „Social force model for pedestrian dynamics“. In: *Physical review E* 51.5 (1995), p. 4282 (cit. on p. 13).

[68] Xiaoyan Hong, Mario Gerla, Guangyu Pei, and Ching-Chuan Chiang. „A group mobility model for ad hoc wireless networks“. In: *Proceedings of the 2nd ACM international workshop on Modeling, analysis and simulation of wireless and mobile systems*. ACM. 1999, pp. 53–60 (cit. on p. 68).

[69] Simon Hudson and Rupert Hudson. „Engaging with consumers using social media: a case study of music festivals“. In: *International Journal of Event and Festival Management* 4.3 (2013), pp. 206–223 (cit. on p. 145).

[70] Pan Hui, Eiko Yoneki, Shu Yan Chan, and Jon Crowcroft. „Distributed community detection in delay tolerant networks“. In: *Proceedings of 2nd ACM/IEEE international workshop on Mobility in the evolving internet architecture*. ACM. 2007, p. 7 (cit. on p. 62).

[71] Hayley Hung, Gwenn Englebienne, and Jeroen Kools. „Classifying social actions with a single accelerometer“. In: *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. ACM. 2013, pp. 207–210 (cit. on p. 146).

[72] Hayley Hung, Gwenn Englebienne, and Laura Cabrera-Quiros. „Detecting conversing groups with a single worn accelerometer“. In: *Proceedings of the 16th International Conference on Multimodal Interaction*. ACM. 2014, pp. 84–91 (cit. on p. 146).

[73] Faisel T Illiyas, Shibu K Mani, AP Pradeepkumar, and Keshav Mohan. „Human stampedes during religious festivals: A comparative review of mass gathering emergencies in India“. In: *International Journal of Disaster Risk Reduction* 5 (2013), pp. 10–18 (cit. on p. 14).

[74] Lorenzo Isella, Juliette Stehlé, Alain Barrat, et al. „What's in a crowd? Analysis of face-to-face behavioral networks“. In: *Journal of theoretical biology* 271.1 (2011), pp. 166–180 (cit. on pp. 42, 48, 59, 80).

[75] Petra Isenberg, Niklas Elmqvist, Jean Scholtz, et al. „Collaborative visualization: definition, challenges, and research agenda“. In: *Information Visualization* 10.4 (2011), pp. 310–326 (cit. on p. 35).

[76] Z Jiangy, W Xiy, Xiang-Yang Li, J Zhaoy, and J Hany. *HiLoc: A TDoA-fingerprint hybrid indoor localization system*. Tech. rep. Technical report, Microsoft Indoor Localization Competition, 2014 (cit. on p. 90).

[77] Naim Kapucu and Vener Garayev. „Collaborative decision-making in emergency and disaster management“. In: *International Journal of Public Administration* 34.6 (2011), pp. 366–375 (cit. on p. 36).

[78] Volker Kirchberg and Martin Tröndle. „The museum experience: Mapping the experience of fine art“. In: *Curator: The Museum Journal* 58.2 (2015), pp. 169–193 (cit. on pp. 91, 118, 125, 141).

[79] Mikkel Baun Kjærgaard, Martin Wirz, Daniel Roggen, and Gerhard Tröster. „Mobile sensing of pedestrian flocks in indoor environments using wifi signals“. In: *Pervasive Computing and Communications (PerCom), 2012 IEEE International Conference on*. IEEE. 2012, pp. 95–102 (cit. on p. 59).

[80] Andreas Köpke, Michael Swigulski, Karl Wessel, et al. „Simulating wireless and mobile networks in OMNeT++ the MiXiM vision“. In: *Proceedings of the 1st international conference on Simulation tools and techniques for communications, networks and systems & workshops*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering). 2008, p. 71 (cit. on p. 70).

[81] Barbara Krausz and Christian Bauckhage. „Loveparade 2010: Automatic video analysis of a crowd disaster“. In: *Computer Vision and Image Understanding* 116.3 (2012), pp. 307–319 (cit. on p. 125).

[82] Tsvi Kuflik, Zvi Boger, and Massimo Zancanaro. „Analysis and prediction of museum visitors' behavioral pattern types“. In: *Ubiquitous Display Environments*. Springer, 2012, pp. 161–176 (cit. on p. 84).

[83] Jennifer R Kwapisz, Gary M Weiss, and Samuel A Moore. „Activity recognition using cell phone accelerometers“. In: *ACM SigKDD Explorations Newsletter* 12.2 (2011), pp. 74–82 (cit. on pp. 13, 146).

[84] Nicholas D Lane, Emiliano Miluzzo, Hong Lu, et al. „A survey of mobile phone sensing“. In: *Communications Magazine, IEEE* 48.9 (2010), pp. 140–150 (cit. on p. 14).

[85] Joel Lanir, Peter Bak, and Tsvi Kuflik. „Visualizing Proximity-Based Spatiotemporal Behavior of Museum Visitors using Tangram Diagrams“. In: *Computer Graphics Forum*. Vol. 33. 3. Wiley Online Library. 2014, pp. 261–270 (cit. on pp. 102, 124).

[86] Celine Latulipe, Erin A Carroll, and Danielle Lottridge. „Love, hate, arousal and engagement: exploring audience responses to performing arts“. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. 2011, pp. 1845–1854 (cit. on p. 146).

[87] Patrick Lazik and Anthony Rowe. „Indoor pseudo-ranging of mobile devices using ultrasonic chirps“. In: *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*. ACM. 2012, pp. 99–112 (cit. on p. 90).

[88] Anna Leask, Ahmed Hassanien, and Philip C Rothschild. „Social media use in sports and entertainment venues“. In: *International Journal of Event and Festival Management* 2.2 (2011), pp. 139–150 (cit. on p. 145).

[89] Nancy G Leveson. „Applying systems thinking to analyze and learn from events“. In: *Safety Science* 49.1 (2011), pp. 55–64 (cit. on p. 35).

[90] Chin-Lung Li, Christos Laoudias, George Larkou, et al. „Indoor geolocation on multi-sensor smartphones“. In: *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*. ACM. 2013, pp. 503–504 (cit. on pp. 90, 125).

[91] Jie Li, Huib de Ridder, APOS Vermeeren, Claudine Conrado, and Claudio Martella. „Designing for crowd well-being: Current designs, strategies and future design suggestions“. In: *IASDR 2013: Proceedings of the 5th International Congress of International Association of Societies of Design Research" Consilience and Innovation in Design", Tokyo, Japan, 26-30 August 2013*. 2013 (cit. on p. 36).

[92] Konrad Lorincz, David J Malan, Thaddeus RF Fulford-Jones, et al. „Sensor networks for emergency response: challenges and opportunities“. In: *Pervasive Computing, IEEE* 3.4 (2004), pp. 16–23 (cit. on p. 14).

[93] Paul Luff and Christian Heath. „The collaborative production of computer commands in command and control“. In: *International Journal of Human-Computer Studies* 52.4 (2000), pp. 669–699 (cit. on p. 14).

[94] Dimitrios Lymberopoulos, Jie Liu, Xue Yang, et al. „A realistic evaluation and comparison of indoor location technologies: Experiences and lessons learned“. In: *Proceedings of the 14th international conference on information processing in sensor networks*. ACM. 2015, pp. 178–189 (cit. on pp. 89, 125, 126).

[95] Wendy E Mackay, Anne-Laure Fayard, Laurent Frobert, and Lionel Médini. „Reinventing the familiar: exploring an augmented reality design space for air traffic control“. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press/Addison-Wesley Publishing Co. 1998, pp. 558–565 (cit. on p. 14).

[96] Oded Maimon and Lior Rokach. *Data mining and knowledge discovery handbook*. Vol. 2. Springer, 2005 (cit. on p. 110).

[97] Claudio Martella, Maarten Van Steen, A Halteren, Claudine Conrado, and Jie Li. „Crowd textures as proximity graphs“. In: *Communications Magazine, IEEE* 52.1 (2014), pp. 114–121 (cit. on pp. 13, 61, 91, 141, 159).

[98] Claudio Martella, Matthew Dobson, Aart Van Halteren, and Maarten Van Steen. „From proximity sensing to spatio-temporal social graphs“. In: *Pervasive Computing and Communications (PerCom), 2014 IEEE International Conference on*. IEEE. 2014, pp. 78–87 (cit. on pp. 54, 95, 97, 131, 160).

[99] Claudio Martella, Ekin Gedik, Laura Cabrera-Quiros, Gwenn Englebienne, and Hayley Hung. „How Was It?: Exploiting Smartphone Sensing to Measure Implicit Audience Responses to Live Performances“. In: *Proceedings of the 23rd ACM international conference on Multimedia*. ACM. 2015, pp. 201–210 (cit. on pp. 118, 141).

[100] Claudio Martella, Armando Miraglia, Marco Cattani, and Maarten van Steen. „Leveraging proximity sensing to mine the behavior of museum visitors“. In: *2016 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE. 2016, pp. 1–9 (cit. on p. 130).

[101] Mauro Martino, Francesco Calabrese, Giusy Di Lorenzo, et al. „Ocean of information: fusing aggregate & individual dynamics for metropolitan analysis“. In: *Proceedings of the 15th international conference on Intelligent user interfaces*. ACM. 2010, pp. 357–360 (cit. on p. 35).

[102] Aleksandar Matic, Venet Osmani, Alban Maxhuni, and Oscar Mayora. „Multi-modal mobile sensing of social interactions“. In: *2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*. IEEE. 2012, pp. 105–114 (cit. on p. 146).

[103] Ramin Mehran, Akira Oyama, and Mubarak Shah. „Abnormal crowd behavior detection using social force model“. In: *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. IEEE. 2009, pp. 935–942 (cit. on p. 125).

[104] Eve Mitleton-Kelly, Ivan Deschenaux, Christian Maag, Matthew Fullerton, and Nihan Celikkaya. *Enhancing Crowd Evacuation and Traffic Management Through AmI Technologies: A Review of the Literature*. Springer, 2013 (cit. on p. 14).

[105] 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 (cit. on p. 126).

[106] Heather L O’Brien and Elaine G Toms. „The development and evaluation of a survey to measure user engagement“. In: *Journal of the American Society for Information Science and Technology* 61.1 (2010), pp. 501–769 (cit. on p. 148).

[107] Sung Park, Ivo Locher, Andreas Savvides, et al. „Design of a wearable sensor badge for smart kindergarten“. In: *Wearable Computers, 2002.(ISWC 2002). Proceedings. Sixth International Symposium on*. IEEE. 2002, pp. 231–238 (cit. on p. 81).

[108] Neal Patwari, Joshua N Ash, Spyros Kyperountas, et al. „Locating the nodes: cooperative localization in wireless sensor networks“. In: *Signal Processing Magazine, IEEE* 22.4 (2005), pp. 54–69 (cit. on p. 126).

[109] Guangyu Pei and Thomas R Henderson. „Validation of OFDM error rate model in ns-3“. In: *Boeing Research Technology* (2010), pp. 1–15 (cit. on p. 72).

[110] Alex Pentland, Tanzeem Choudhury, Nathan Eagle, and Push Singh. „Human dynamics: computation for organizations“. In: *Pattern Recognition Letters* 26.4 (2005), pp. 503–511 (cit. on pp. 59, 159).

[111] Ronald W Perry. „Incident management systems in disaster management“. In: *Disaster Prevention and Management: An International Journal* 12.5 (2003), pp. 405–412 (cit. on p. 14).

[112] John W Ratcliff and David E Metzener. „Pattern-matching-the gestalt approach“. In: *Dr Dobbs Journal* 13.7 (1988), p. 46 (cit. on p. 98).

[113] Matthew Reason and Dee Reynolds. „Kinesthesia, empathy, and related pleasures: An inquiry into audience experiences of watching dance“. In: *Dance Research Journal* 42.02 (2010), pp. 491–775 (cit. on p. 145).

[114] Madhu C Reddy, Sharoda A Paul, Joanna Abraham, et al. „Challenges to effective crisis management: using information and communication technologies to coordinate emergency medical services and emergency department teams“. In: *International journal of medical informatics* 78.4 (2009), pp. 259–269 (cit. on p. 14).

[115] Sasank Reddy, Min Mun, Jeff Burke, et al. „Using mobile phones to determine transportation modes“. In: *ACM Transactions on Sensor Networks (TOSN)* 6.2 (2010), p. 13 (cit. on p. 146).

[116] Jörg Reichardt and Stefan Bornholdt. „Statistical mechanics of community detection“. In: *Physical Review E* 74.1 (2006), p. 016110 (cit. on p. 53).

[117] Stephen Reicher. „The psychology of crowd dynamics“. In: *Blackwell handbook of social psychology: Group processes* (2001), pp. 182–208 (cit. on p. 13).

[118] Rönne Reimann, Arne Bestmann, and Mirjam Ernst. „Locating technology for AAL applications with direction finding and distance measurement by narrow bandwidth phase analysis“. In: *International Competition on Evaluating AAL Systems through Competitive Benchmarking*. Springer. 2012, pp. 52–62 (cit. on pp. 90, 125).

[119] Craig W Reynolds. „Flocks, herds and schools: A distributed behavioral model“. In: *ACM SIGGRAPH computer graphics* 21.4 (1987), pp. 25–34 (cit. on p. 45).

[120] Daniel Roggen, Martin Wirz, Gerhard Tröster, and Dirk Helbing. „Recognition of crowd behavior from mobile sensors with pattern analysis and graph clustering methods“. In: *Networks and Heterogeneous Media,* (2011) (cit. on pp. 13, 41, 42, 54).

[121] Elizabeth B-N Sanders and Pieter Jan Stappers. *Convivial toolbox: Generative research for the front end of design*. BIS, 2012 (cit. on p. 17).

[122] Johan Schubert and Robert Suzic. „Decision support for crowd control: Using genetic algorithms with simulation to learn control strategies“. In: *Military Communications Conference, 2007. MILCOM 2007. IEEE*. IEEE. 2007, pp. 1–7 (cit. on p. 36).

[123] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. „The experience of presence: Factor analytic insights“. In: *Presence* 10.3 (2001), pp. 266–281 (cit. on p. 148).

[124] Eric WK See-To, Savvas Papagiannidis, and Vincent Cho. „User experience on mobile video appreciation: How to engross users and to enhance their enjoyment in watching mobile video clips“. In: *Technological Forecasting and Social Change* 79.8 (2012), pp. 1484–1494 (cit. on p. 148).

[125] AA Siddiqui and SMV Gwynne. „Employing pedestrian observations in engineering analysis“. In: *Safety science* 50.3 (2012), pp. 478–493 (cit. on p. 13).

[126] Jonathan D Sime. „Crowd psychology and engineering“. In: *Safety science* 21.1 (1995), pp. 1–14 (cit. on p. 13).

[127] RA Smith. „Density, velocity and flow relationships for closely packed crowds“. In: *Safety science* 18.4 (1995), pp. 321–327 (cit. on p. 25).

[128] Bi Song, Ricky J Sethi, and Amit K Roy-Chowdhury. „Wide area tracking in single and multiple views“. In: *Visual Analysis of Humans*. Springer, 2011, pp. 91–107 (cit. on p. 125).

[129] RJC Stanton and GK Wanless. „Pedestrian movement“. In: *Safety science* 18.4 (1995), pp. 291–300 (cit. on p. 25).

[130] Bernhard Steffen and Armin Seyfried. „Methods for measuring pedestrian density, flow, speed and direction with minimal scatter“. In: *Physica A: Statistical mechanics and its applications* 389.9 (2010), pp. 1902–1910 (cit. on p. 125).

[131] Catherine Stevens, Heather Winskel, Clare Howell, et al. „Direct and indirect methods for measuring audience reactions to contemporary dance“. In: (2009) (cit. on p. 146).

[132] G Keith Still. „Crowd dynamics“. PhD thesis. University of Warwick, 2000 (cit. on p. 13).

[133] Robert Strohmaier, Gerhard Sprung, Alexander Nischelwitzer, and Sandra Schadenbauer. „Using Visitor Flow Visualization to Improve Visitor Experience in Museums and Exhibitions“. In: (2015) (cit. on p. 124).

[134] Lucy Suchman. „Centers of coordination: A case and some themes“. In: *Discourse, Tools and Reasoning*. Springer, 1997, pp. 41–62 (cit. on p. 14).

[135] Latanya Sweeney. „k-anonymity: A model for protecting privacy“. In: *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10.05 (2002), pp. 557–570 (cit. on p. 37).

[136] Martin Tröndle, Steven Greenwood, Konrad Bitterli, and Karen van den Berg. „The effects of curatorial arrangements“. In: *Museum Management and Curatorship* 29.2 (2014), pp. 140–173 (cit. on p. 105).

[137] *USA Today. Providence theater experiments with 'tweet seats'*. `http://www.usatoday.com/story/tech/2013/01/27/theater-tweet-seats/1868693/`. 2013 (cit. on p. 145).

[138] Laurens Van der Maaten and Geoffrey Hinton. „Visualizing Data using t-SNE.“ In: *Journal of Machine Learning Research* 9.11 (2008) (cit. on p. 155).

[139] Fiammetta Venuti, Luca Bruno, and Nicola Bellomo. „Crowd dynamics on a moving platform: Mathematical modelling and application to lively footbridges“. In: *Mathematical and Computer Modelling* 45.3 (2007), pp. 252–269 (cit. on p. 13).

[140] Eliséo Véron and Martine Levasseur. *Ethnographie de l'exposition: l'espace, le corps et le sens*. Centre Georges Pompidou, Bibliothèque publique d'information, 1989 (cit. on p. 84).

[141] Alessandro Vinciarelli, Maja Pantic, and Hervé Bourlard. „Social signal processing: Survey of an emerging domain“. In: *Image and Vision Computing* 27.12 (2009), pp. 1743–1759 (cit. on p. 13).

[142] Spyros Voulgaris, Matthew Dobson, and Maarten Van Steen. „Decentralized Network-level Synchronization in Mobile Ad Hoc Networks“. In: *ACM Transactions on Sensor Networks (TOSN)* 12.1 (2016), p. 5 (cit. on pp. 87, 148).

[143] Chen Wang and Pablo Cesar. „Do we react in the same manner?: comparing GSR patterns across scenarios“. In: *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*. ACM. 2014, pp. 501–510 (cit. on p. 156).

[144] Chen Wang, Erik N Geelhoed, Phil P Stenton, and Pablo Cesar. „Sensing a live audience“. In: *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM. 2014, pp. 1909–1912 (cit. on p. 146).

[145] Joe H Ward Jr. „Hierarchical grouping to optimize an objective function“. In: *Journal of the American statistical association* 58.301 (1963) (cit. on p. 110).

[146] Elias Weingartner, Hendrik Vom Lehn, and Klaus Wehrle. „A performance comparison of recent network simulators“. In: *2009 IEEE International Conference on Communications*. IEEE. 2009, pp. 1–5 (cit. on p. 68).

[147] Ferdinanda Elfrida Hubertina Wijermans. *Understanding crowd behaviour: Simulating situated individuals*. University of Groningen Groningen, 2011 (cit. on pp. 13, 16).

[148] Nanda Wijermans, Claudine Conrado, Maarten van Steen, Claudio Martella, and Jie Li. „A landscape of crowd-management support: An integrative approach“. In: *Safety science* 86 (2016), pp. 142–164 (cit. on pp. 11, 125, 141).

[149] Gillian Wilson. „Multimedia tour programme at tate modern“. In: *Museums and the Web*. Vol. 3. 2004 (cit. on pp. 91, 125).

[150] Moe Z Win, Andrea Conti, Santiago Mazuelas, et al. „Network localization and navigation via cooperation“. In: *Communications Magazine, IEEE* 49.5 (2011), pp. 56–62 (cit. on p. 126).

[151] William E Winkler. „String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage.“ In: (1990) (cit. on pp. 98, 110).

[152] Martin Wirz, Tobias Franke, Eve Mitleton-Kelly, et al. „Coenosense: A framework for real-time detection and visualization of collective behaviors in human crowds by tracking mobile devices“. In: *Proceedings of the European Conference on Complex Systems 2012*. Springer. 2013, pp. 353–361 (cit. on p. 35).

[153] Martin Wirz, Daniel Roggen, and Gerhard Troster. „Decentralized detection of group formations from wearable acceleration sensors“. In: *Computational Science and Engineering, 2009. CSE'09. International Conference on*. Vol. 4. IEEE. 2009, pp. 952–959 (cit. on p. 13).

[154] Martin Wirz, Tobias Franke, Daniel Roggen, et al. „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*. IEEE. 2012, pp. 367–372 (cit. on p. 13).

[155] Martin Wirz, Daniel Roggen, and G Troster. „User acceptance study of a mobile system for assistance during emergency situations at large-scale events". In: *Human-Centric Computing (HumanCom), 2010 3rd International Conference on*. IEEE. 2010, pp. 1–6 (cit. on p. 35).

[156] Martin Wirz, Eve Mitleton-Kelly, Tobias Franke, et al. *Using mobile technology and a participatory sensing approach for crowd monitoring and management during large-scale mass gatherings*. Springer, 2013 (cit. on p. 13).

[157] Henk Wymeersch, Jaime Lien, and Moe Z Win. „Cooperative localization in wireless networks". In: *Proceedings of the IEEE* 97.2 (2009), pp. 427–450 (cit. on p. 126).

[158] Zhongyang Xiong, Ruotian Chen, Yufang Zhang, and Xuan Zhang. „Multi-density dbscan algorithm based on density levels partitioning". In: *Journal of Information and Computational Science* 9.10 (2012), pp. 2739–2749 (cit. on p. 66).

[159] Kazuo Yano, Koji Ara, Norihiko Moriwaki, and Hiroyuki Kuriyama. „Measurement of human behavior: creating a society for discovering opportunities". In: *Hitachi Review* 58.4 (2009), p. 139 (cit. on pp. 59, 80).

[160] Andrew C Yao. „Protocols for secure computations". In: *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*. IEEE. 1982, pp. 160–164 (cit. on p. 37).

[161] S Yaseen, A Al-Habaibeh, D Su, and F Otham. „Real-time crowd density mapping using a novel sensory fusion model of infrared and visual systems". In: *Safety science* 57 (2013), pp. 313–325 (cit. on pp. 13, 125).

[162] Yuji Yoshimura, Fabien Girardin, Juan Pablo Carrascal, Carlo Ratti, and Josep Blat. *New tools for studying visitor behaviours in museums: a case study at the Louvre*. Springer-Verlag, 2012 (cit. on pp. 91, 125).

[163] Yaoxuan Yuan, Jizhong Zhao, Chen Qiu, and Wei Xi. „Estimating crowd density in an RF-based dynamic environment". In: *IEEE Sensors Journal* 13.10 (2013), pp. 3837–3845 (cit. on p. 126).

[164] Marco Zúñiga Zamalloa and Bhaskar Krishnamachari. „An analysis of unreliability and asymmetry in low-power wireless links". In: *ACM Transactions on Sensor Networks (TOSN)* 3.2 (2007), p. 7 (cit. on p. 72).

[165] Nikos Zarboutis and Nicolas Marmaras. „Design of formative evacuation plans using agent-based simulation“. In: *Safety Science* 45.9 (2007), pp. 920–940 (cit. on p. 13).

[166] Beibei Zhan, Dorothy N Monekosso, Paolo Remagnino, Sergio A Velastin, and Li-Qun Xu. „Crowd analysis: a survey“. In: *Machine Vision and Applications* 19.5-6 (2008), pp. 345–357 (cit. on pp. 13, 125).

[167] Sheng Zhang, Weihong Wang, James Ford, and Fillia Makedon. „Learning from Incomplete Ratings Using Non-negative Matrix Factorization.“ In: *SDM*. Vol. 6. SIAM. 2006, pp. 548–552 (cit. on p. 115).

[168] Tong Zhang, Jue Wang, Ping Liu, and Jing Hou. „Fall detection by embedding an accelerometer in cellphone and using KFD algorithm“. In: *International Journal of Computer Science and Network Security* 6.10 (2006), pp. 277–284 (cit. on p. 146).

[169] Gang Zhou, Tian He, Sudha Krishnamurthy, and John A Stankovic. „Impact of radio irregularity on wireless sensor networks“. In: *Proceedings of the 2nd international conference on Mobile systems, applications, and services*. ACM. 2004, pp. 125–138 (cit. on pp. 70, 92).

[170] Han Zou, Lihua Xie, Qing-Shan Jia, and Hengtao Wang. „Platform and algorithm development for a rfid-based indoor positioning system“. In: *Unmanned Systems* 2.03 (2014), pp. 279–291 (cit. on p. 90).

List of Figures

List of Tables

Summary

We are witnessing massive volumes of data about our behavior being produced and collected on a daily basis. This data comprises information about our multitude of activities online, including the World Wide Web and social networking and social media sites, as well a offline, recorded through the myriad of sensors installed in our smart phones and smart watches. Effective modeling and analysis of such data opens the doors to a number of applications and services that are more adaptive and personalized to the user, and to new paradigms like the Internet of Things.

Spatio-temporal proximity is one of the signals that can be captured with existing technology like, for example, radio-based proximity sensors (e.g., Bluetooth Low Energy transceivers in modern smart phones). Spatio-temporal proximity data can be the foundation of a number of analyses of social behavior and in particular, the focus of this dissertation, of crowd behavior. Crowd behavior has a strong spatio-temporal nature, and it lends itself to be studied from this perspective. Furthermore, a crowd is collective in nature, and the behavior of the members of the crowd cannot be understood when looking at the individuals in isolation, but only when the behavior of all the members, together with the relationships, is analyzed. We coined the term “crowd texture” to refer to the interleaved spatio-temporal relationships between the member of a crowd that characterize specific crowd behaviors.

We started our work by interviewing ten crowd managers to understand the current crowd management practices and use of technology. We have found that crowd managers are interested in particular in means of increasing situation awareness, prediction, and intervention, in particular regarding movements and density variations in the crowd (e.g., flows and congestions), but also higher level aspects such as mood, in particular when conditions are critical. We have also found that crowd managers rely on technology only to a limited extent, as the current

existing solutions are found unreliable in most cases. Based on these results, we set as the main research question whether we can leverage spatio-temporal proximity information collected through proximity sensors to gain a better understanding of the behavior of a crowd, to support the work of researchers and practitioners towards safer and more comfortable crowds.

In this dissertation, we model spatio-temporal proximity as a series of proximity graphs, where each graph represents which entities were within a certain range of distance at a specific moment in time. In a proximity graph a vertex represents an entity while an edge represents a positive proximity relationship between two entities. By mining a series of proximity graphs it is possible to identify and quantify different crowd behavior and dynamics, like pedestrian lanes, clogging, social groups, as well as group behavior. Together with the model, we propose a general approach to sense and collect the texture of a crowd though radio-based proximity sensors.

Due to the inherent limitations of wireless sensors networks and radio-based communication, proximity graphs collected with proximity sensors are often noisy and incomplete, meaning that proximity relationships can be wrongly added or missing. Proximity graphs can be filtered to infer the missing relationships and remove the noisy ones. In particular, we present a novel filter based on a density-based clustering algorithm that identifies automatically intervals of proximity between entities in the data, and nearly doubles the accuracy of the measurement. We also present a solution to compute optimal parameters for the technique and we validate it both in simulation and through real-world experiments.

Once proximity data is collected and filtered, it can be analyzed to identify group behavior and gain insight about the behavior of a crowd. We have deployed our infrastructure inside of the CoBrA museum of Amsterdam to track the behavior of some of its visitors (who volunteered to wear one of our sensors), by installing some proximity sensors at exhibits. Working together with the museum staff, we have produced a number of visualizations that helped the museum staff to better understand the amount of time spent by the visitors at each exhibit and the paths within the exhibition rooms followed by them, to name a few examples.

Crowd density is one of the characteristics crowd managers are most interested in, as it can be the cause (or fingerprint) of extreme and dangerous situations. Crowd density is also a major obstacle to radio-based communication, hindering in particular proximity sensing systems that rely on long-range transmissions between wearable sensors and fixed infrastructure. This means that these solutions often fail in the very conditions where they are most needed. We present a solution that leverages mobile-to-mobile proximity sensing to overcome limitations of mobile-to-fixed solutions when crowd density is high. We deployed our solution at the NEMO museum in Amsterdam, a museum located in a complex building characterized by a multi story open space that is particularly challenging for radio-based communication. During our experiment, crowd density at peak was high, with more than 600 visitors wearing our sensor (and overall around 2400 visitors in the museum at that moment). We show that crowd density can be leveraged to increase measurement accuracy as well as to gain insights about the behavior of the visitors.

Spatio-temporal proximity is an effective proxy for measuring crowd behavior, yet it cannot be used to understand higher-level aspects of crowd behavior, which are important to study, for example, mood. In fact, other signals like galvanic skin response, audio and acceleration are more suitable to study the response and the inner experience of the members of the crowd. Still, spatio-temporal proximity can play an important role in these studies, as it can help the understanding of how the response of the crowd (e.g., mood and emotions) spread spatio-temporally across the crowd. We have deployed our proximity sensing infrastructure together with accelerometers that we have used to predict the response of the audience of a dance performance. We show that by measuring proximity and movement of the audience during the performance as well as during mingling sessions before and after the performance, we can predict the enjoyment of the audience.

The work presented in this dissertation shows promising results of the application of pervasive and ubiquitous proximity sensing to the study of social behavior that open a number of new research directions for future work.

Samenvatting

We zien dat er tegenwoordig op dagelijkse basis enorm veel gegevens over ons gedrag verzameld wordt. Dit zijn gegevens over onze online activiteiten, zoals surfgedrag, sociale netwerken en sociale media sites, maar ook offline activiteiten, die door allerlei sensoren wordt opgenomen in onze telefoons en horloges. We kunnen applicaties en diensten beter op de gebruiker afgestemd maken als we deze gegevens goed kunnen modelleren en analyseren. Dan kan ook het meeste uit nieuwe werkwijzen zoals het Internet of Things gehaald worden.

Nabijheid in tijd en ruimte is een van de signalen die met huidige technologie al opgenomen kunnen worden, zoals radio-gebaseerde nabijheidssensoren (bijv. Bluetooth Low Energy radio chips in moderne telefoons). Analyse van deze nabijheids informatie kan de hoeksteen vormen van een aantal analyses van sociaal gedrag, en met name het gedrag van een menigte. Dit is het aandachtsgebied van dit proefschrift. Het gedrag van mensenmassa's kenmerkt zich door eigenschappen die zich in tijd en ruimte uitdrukken. Deze eigenschappen lenen zich goed voor studie vanuit dit perspectief. Verder is een menigte van nature een gezamenlijk fenomeen. Gedrag van individuen in de menigte kan niet begrepen worden door naar de geïsoleerde individuen te kijken, maar alleen door naar het gedrag van alle individuen en de relaties met elkaar te kijken en te analyseren. We stellen de term "menigte structuur" voor voor het beschrijven van de relaties in tijd en ruimte tussen individuen in een menigte die specifiek gedrag in menigtes voortbrengt.

We zijn begonnen met tien menigte beheerders te ondervragen naar huidige beheerstechnieken en technologie gebruik. We zijn erachter gekomen dat menigte beheerders met name geïnteresseerd zijn in het verhogen van de situatie bewustheid, voorspelling, en interventie, met name wat betreft bewegingen en dichtheid variaties in de menigte (bijv. stromen en opstoppingen), maar ook hoger-niveau aspecten zoals stemming, voornamelijk wanneer omstandigheden kritiek zijn. We zijn er ook achter gekomen dat menigte beheerders slechts in beperkte mate zich van technologie afhankelijk maken, omdat huidige oplossingen in de meeste gevallen te onbetrouwbaar zijn gebleken. Hierop bouwend hebben we ons de onderzoeksvraag gesteld: kunnen we door nabijheidssensoren opgevangen informatie, oftewel nabijheidsinformatie in ruimtelijke- en tijds-zin, om beter begrip te krijgen van het gedrag van een menigte, om het werk van onderzoekers en beoefenaars te ondersteunen om veiligere en prettigere menigtes te bewerkstelligen.

In dit proefschrift modelleren we nabijheid met een serie van nabijheidsgrafen, waarbij elke graaf weergeeft welke entiteiten zich op een bepaald moment binnen een zekere afstand van elkaar bevonden. In een nabijheidsgraaf is een punt (of knoop) een entiteit, en een lijn (of kant) een nabijheidsrelatie tussen twee entiteiten. We kunnen een aantal nabijheidsgrafen analyseren, en daarmee een verschillende gevallen van menigte gedrag en dynamiek identificeren en kwantificeren. Voorbeelden zijn voetgangerspaden, opstoppingen, sociale groepen, en groepsgedrag. Samen met dit model stellen we een algemene aanpak van het meten en verzamelen van de structuur van een menigte door middel van radiosensoren voor.

Vanwege inherente beperkingen aan draadloze sensornetwerken en radiocommunicatie zijn nabijheidsgrafen die verzameld zijn met nabijheidssensoren niet compleet. Dit houdt in dat nabijheidsrelaties onterecht aanwezig zijn of onterecht afwezig zijn. Nabijheidsrelaties kunnen gefilterd worden om de missende relaties aan te vullen en de onterechte weg te halen. We presenteren een automatisch filter dat gebaseerd is op een clustering algoritme dat automatisch intervallen van nabijheid tussen entiteiten identificeert in de metingen. Deze techniek maakt de nauwkeurigheid van de meting bijna twee keer zo groot. We presenteren ook een techniek om de optimale parameters voor de techniek te berekenen. Deze worden zowel door simulatie als door echte experimenten solide getoond.

Zodra nabijheidsdata gefilterd en verzameld is kan het geanalyseerd worden om groepsgedrag te vinden en inzicht te krijgen in het gedrag van een menigte. We hebben onze infrastructuur in het CoBrA museum van Amsterdam uitgerold om het gedrag van een aantal van de bezoekers te volgen (die vrijwillig een van onze sensoren gedragen hebben), door wat nabijheidssensoren bij tentoonstellingen te installeren. We hebben samen met het personeel van het museum wat visualisaties gemaakt waardoor het gedrag van de bezoekers beter begrepen kon worden - bij voorbeeld hoe lang bezoekers bij elk onderdeel van de tentoonstelling stil hebben gestaan, en de paden die binnen de ruimtes gevolgd zijn.

Menigte dichtheid is een van de eigenschappen waar menigte beheerders het meeste in geinteresseerd zijn, omdat het de aanleiding (of kenmerk) van extreme of gevaarlijke situaties kan zijn. Menigte dichtheid is ook een groot obstakel in radio-gebaseerde communicatie, met name omdat nabijheids detectie systemen gehinderd worden die afhankelijk zijn van lange-afstandscommunicatie met vaste infrastructuur. Dit betekent dat deze oplossingen vaak falen in precies die omstandigheden waarin ze het meest nodig zijn. We presenteren een oplossing die mobiel-naar-mobiel nabijheidsdetectie inzet om de beperkingen van mobiel-naar-vaste oplossingen te overwinnen. We hebben onze oplossing bij het NEMO museum in Amsterdam uitgerold, een museum dat in een complex gebouw gehuisvest is dat gekarakteriseerd wordt door een open ruimte van meerdere verdiepingen dat bij uitstek problematisch is voor radiocommunicatie. Tijdens ons experiment was de menigte dichtheid op piekmomenten hoog, met meer dan 600 bezoekers die onze sensor droegen (van een totaal van 2400 bezoekers in het museum op dat moment). We laten zien dat menigte dichtheid juist ingezet kan worden om zowel meting nauwkeurigheid te verhogen als inzicht te krijgen in het gedrag van de bezoekers.

Ruimte-tijds nabijheid is een effectief surrogaat voor het meten van menigte gedrag, maar het kan niet gebruikt worden om hoger-niveau aspecten van menigte gedrag te begrijpen, die ook belangrijk zijn om te bestuderen, zoals stemming. Andere metingen zoals elektrische weerstand van de huid, audio, en acceleratie zijn bruikbaarder om de respons en innerlijke ervaring van de leden van de massa te bestuderen. Ruimtetijds nabijheid kan echter toch een belangrijke rol spelen in deze studies, omdat het de reactie van de menigte (bijv. Stemming en emoties) en de ruimte-tijds verspreiding door de massa heen kan helpen begrijpen. We hebben onze nabijheids detectie infrastructuur samen met acceleratiemeters uitgerold om de reactie van het publiek van een dansuitvoering te voorspellen. We laten zien dat door nabijheid en beweging van het publiek zowel tijdens de uitvoering als tijdens de sociale interactie voor en na de uitvoering te meten, we de plezierbeleving van het publiek kunnen voorspellen.

Het werk dat in dit proefschrift wordt gepresenteerd laat veelbelovende resultaten zien van de toepassing van alom aanwezige nabijheidsdetectie op de studie van sociaal gedrag dat een aantal nieuwe onderzoeksrichtingen voor verder onderzoekswerk opent.