FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Using Ubisense to Model Human Behaviour in Emergency Situations

João Gonçalo Areias Miranda Vasconcelos

Mestrado Integrado em Engenharia Eletrotécnica e de Computadores

Supervisor: Rosaldo J. F. Rossetti

Second Supervisor: Marcelo R. Petry

July 26, 2013

Resumo

Os modelos de comportamento humano são um contributo importante para as ferramentas que ajudam a mapear e arquitectar planos de evacuação eficientes. Estes modelos precisam de dados que podem ser fornecidos por sistemas de localização como o Ubisense.

Este estudo tem como objectivo explorar o uso do sistema de localização em tempo real Ubisense aquando da modelação do comportamento humano e ainda desenvolver novos conhecimentos sobre padrões de trajectórias de pessoas em evacuações e outras situações de stress.

De forma a colectar dados de movimentações pedestres, concebemos e executamos experiências de comportamento pedonal de pequena escala com voluntários, em diferentes cenários e com o objectivo de replicar várias situações encontradas em instalações reais, nas quais as movimentações de cada participante foram monitoradas e gravadas com recurso ao sistema Ubisense.

Embora as trajetorias recolhidas em estado bruto parecerem imprecisas e irregulares, as tentativas para filtrar e limpar os dados foram bem sucedidas, melhorando-os e preparando-os para a análise posterior. A fim de melhor caracterizar e dar significado aos dados, foram calculados descritores básicos de movimento e exploradas diferentes técnicas de visualização.

Por fim, procurou-se inferir comportamento humano a partir de trajetórias pedestres, dividindo-as em segmentos ou sub-trajetórias que compartilham propriedades semelhantes e agrupando as sub-trajetórias em conjuntos que simbolizavam um determinado comportamento. Seguidamente usaram-se as técnicas de análise anteriormente exploradas para caracterizar estatisticamente cada conjunto e o seu comportamento associado.

Os resultados obtidos demonstraram que a nossa abordagem é válida, pois fomos capazes de extrair padrões e a sua caracterização. Além disso, o uso de tecnologia de monitoramento UWB parece ser uma abordagem viável, pois apesar de devido a limitações como a falta de fina precisão espacial não a tornem ideal para extrair as propriedades microscópicas do tráfego, é particularmente adequada para situações em que a utilização de vídeo é menos aplicável e o posicionamento minucioso não é um problema, como por exemplo na análise macroscópica.

Abstract

The models of human behaviour are an important input for the tools that that help mapping and design efficient evacuation plans. These models must be calibrated and validated with reliable data that can be provided by systems like the Ubisense real time location system.

This study aims to explore the usage of the Ubisense real time location system when modelling human behaviour and develop new insights about patterns of human trajectories in evacuations and other stressful situations.

Towards gathering pedestrian movement data, we devised and performed small scale walking behaviour experiments with volunteers, taking place in different scenarios and aiming to replicate several situations found in real facilities, where the movements of each participant was tracked and recorded with the help of the Ubisense system.

Although the raw gathered trajectories were found to be noisy and imprecise, attempts to filter and clean the data were successful, improving and preparing it for the subsequent analysis. In order to further characterize and make sense of the data, we calculated basic motion descriptors and explored different visualization techniques.

Finally, we attempted to infer human behaviour from pedestrian trajectories by partitioning them into segments or sub-trajectories that shared similar properties and aggregated the sub-trajectories into clusters that symbolized a certain behaviour, and then used the previously explored analysis techniques to statistically characterize each cluster and associated behaviour.

The results we obtained demonstrated that our approach was valid, as we were able to extract patterns and its characterising variables. Also, usage of UWB tracking technology seems a viable approach, as although limitations like the lack of fine spatial precision make it not optimal for extracting microscopic properties of traffic, it is particularly suited for for scenarios where video recording is less applicable and pedestrian fine positioning is not an issue, such as for macroscopic analysis.

Agradecimentos

Este projecto de pesquisa não teria sido possível sem o apoio de algumas pessoas que, directa ou indirectamente, tornaram viável a realização e conclusão da minha tese.

Primeiro e mais que tudo gostaria de agradecer aos meus pais pelo apoio incondicional, paciência e perseverança que têm tido para comigo durante todo o curso e principalmente durante a elaboração da minha tese. Por terem fé no que faço e por toda a compreensão e suporte que mostraram ter em todas as circunstâncias.

O sucesso deste projecto dependeu grandemente do incentivo e directrizes dos meus orientadores: Professor Dr. Rosaldo Rosetti e Eng. Marcelo Petry aos quais gostaria de expressar a minha gratidão. Gostaria também de prestar um sincero agradecimento ao Eng. João Almeida pela sua útil e valiosa assistência, apoio e orientação.

Aproveito esta oportunidade para agradecer à minha irrestrita sequaz Sara Carreira por conseguir na sua maneira infatigável e ininteligível manter-me concentrado nos meus objectivos e principalmente pela sua contínua disposição para me motivar. Finalmente, mas não menos importante gostaria de mencionar a minha gratidão para com meu irmão Diogo que me ajudou em várias e variadas situações durante o decurso deste projecto.

A todos estes um franco ‘obrigado’.

Gonçalo Vasconcelos

Contents

List of Figures

List of Tables

Symbols and Abbreviations

AOA	Angle of Arrival
API	Application Programming Interface
CA	Cellular Automata
CAD	Computer-aided Design
GIS	Geographic Information System
KDE	Kernel Density Estimation
LIACC	Laboratório de Inteligência Artificial e Ciência de Computadores (Artificial Intelligence and Computer Science Laboratory)
MDL	Minimum Description Length
MRD	Maximum Redundant Distance
RF	Radio-Frequency
RFID	Radio-Frequency Identification
RTLS	Real-Time Location System
TDOA	Time-Difference of Arrival
UWB	Ultra-Wideband

Chapter 1

Introduction

This chapter presents the context in which the dissertation was carried out. It begins by introducing the dissertation's theme, motivation, scope and goals. A detailed characterization of the problems to be tackled and corresponding methodological approach follows. Finally the expected contributions are enumerated and the rest of the document structure is summarized.

1.1 Motivation

The last couple of decades have witnessed an increasing number of mass events, in which large numbers of people flock to confined areas. Some of these large events are related to sports, entertainment, culture or religion and are held regularly throughout the world. The safety of people in emergency situations where panic can lead to unexpected or anti-social behaviour has been a subject of interest and study of researchers, engineers and authorities [1]. The study of pedestrian motion and behaviour is thus crucial for planning efficient evacuation strategies to assure safety for people in situations such as these, which can lead to a large number of injured and casualties in an evacuation scenario. However, studying crowd behaviour in emergency situations is challenging since it often requires exposing real people to the actual, possibly dangerous, environments [2]. Furthermore, evacuation exercises are often too expensive and time consuming to be a standard measure for evacuation analysis [3]. Therefore the usage of evacuation simulations built upon models of pedestrian dynamics, have become essential tools in evaluating safety and designing buildings, as well as planning emergency evacuation.

1.2 Scope

This thesis is carried out within the scope of the final dissertation in the Master Course in Electrical and Computer Engineering at Faculdade de Engenharia da Universidade do Porto. It aims to study and improve techniques for tracking pedestrians carrying RFID tags in order to study the movement and behaviour of humans, particularly in situations of panic or emergency. The

resulting data from the movement of several people shall be analysed and filtered to be used in a framework of human behaviour elicitation.

1.3 Problem statement

Agent-based models of pedestrian dynamics play an important role in predicting the dynamical properties of large human crowds and are used as tools for designing buildings and planning emergency egress. These models must, therefore, be calibrated against reliable data to ensure their validity.

To improve the knowledge related to pedestrian dynamics, it is necessary to develop experiments with volunteers where their movements are recorded and data is subsequently mined in order to extract meaningful information.

Data collection on crowds and pedestrians has been traditionally based on direct observation and time-lapse films [4]. Video has become an important tool as several applications for automatic extraction of pedestrian trajectories have been developed [5]. However, the usage of video recordings presents several limitations on the scenarios allowed for pedestrian experiments.

Hopping to widen the breath of possible experiment scenarios, this project will perform data collection using Ubisense, a UWB real-time location system, to track tags carried by pedestrians.

1.4 Aim and goals

This project aims to extract relevant data from pedestrian movement experiments to be used as framework for human behaviour elicitation. In specific, the goals to be achieved are the following:

- Study and improve techniques of tracking human movement using UWB RFID tags;
- Conduct experiments with volunteers and record tracked trajectories;
- Extract relevant information from the recorded trajectories;
- Use data mining and inference techniques to elicitate human behaviour.

1.5 Methodological approach

The methodological approach proposed in this project is formed from three main components. Each of the components constitutes a task itself, which are listed below:

- **Problem analysis** — Perform a literature review to gather insight into the problem at hand;
- **Data collection** — Devise and conduct pedestrian experiments in order to provide relevant data for analysis;
- **Data analysis** — Use behaviour mining techniques to infer pedestrian behaviour from the previously collected data;

1.6 Expected contributions

The usage of ultra-wideband radio frequency based systems for pedestrian experiment data collection allows some advantages over traditional methods such as automatic extraction of trajectories from video recordings. Line-of-sight restrictions are minimized and low ceiling areas present no limitations. It also encompasses a wider breath of simulation scenarios, such as limited visibility situations e.g. dark or smoke filled rooms, which are very common in emergency situations due to fire. Assigning identifiable tags to individual participants allows correlation between characteristics of the individual pedestrian (e.g. gender, height and age) with its trajectory. This might allow better understanding of the dynamics of heterogeneous crowds and study the effect of outliers like elderly people or people with mobility impairments.

The data resulting from this thesis will be used within the context of a larger project: the "mSPEED" Simulator, an integrated framework constituting a unique tool for agent-based "Modelling and Simulation of Pedestrian Emergent Evacuation Dynamics". This tool will be used for validating new and existing building layouts, helping planners to develop and improve emergency plans and safety systems, training occupants using virtual drills, and helping fire-fighters and rescuers to develop plans, rescue strategies, and learn how to deal effectively with crowds during emergencies and critical situations.

1.7 Document structure

Beyond this introduction, this document further comprises four more chapters.

In Chapter 2, related research is described, based on a literature review of relevant topics of pedestrian behaviour and modelling.

Chapter 3, presents the methodological approach on how the problem is to be tackled in this project. Methods, technologies, algorithms and tools used are enumerated and a detailed methodology for the steps that need to be completed to sucessfully achieve the thesis goals are proposed.

The results gathered during the development of the project are presented, analysed and discusses in Chapter 4.

Finally, in Chapter 5 are presented the final conclusions and the directions for further developments within this project's scope are laid out.

Chapter 2

Literature Review

This chapter presents the state-of-the-art of human behaviour in emergency situations and the fundamentals of pedestrian flow theory.

2.1 Introduction

Empirical studies of pedestrians have been carried out for more than four decades [6, 7, 8]. Comprehensive and accurate simulations of emergency situations and evacuation require not only human movement models, but human psychology and behaviour models as well [9].

Hoogendoorn et al. [10] describes pedestrian behaviour at three levels: strategic, tactical and operational. An overview of these decision levels is presented in figure 2.1.

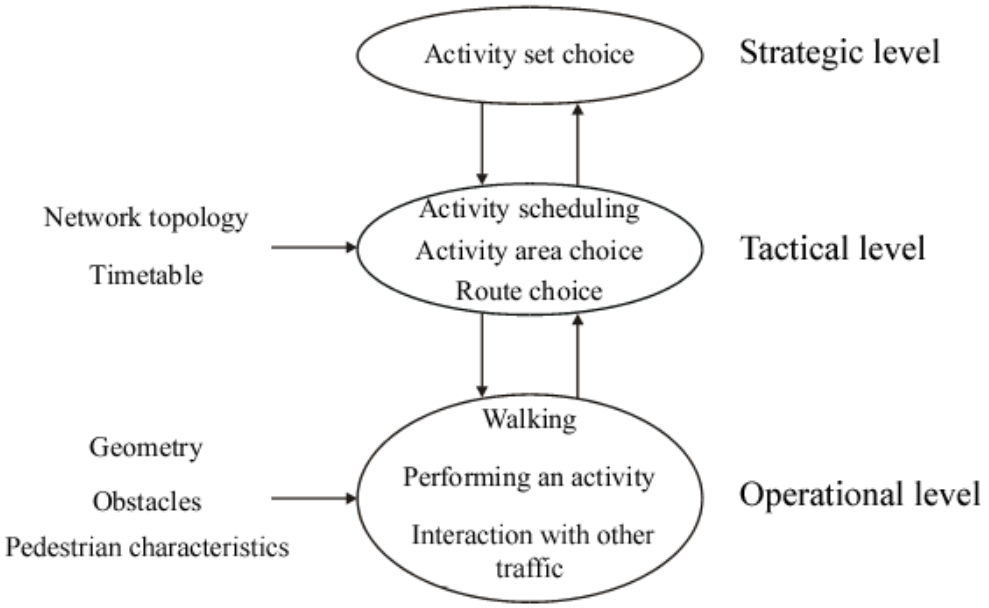


Figure 2.1: Levels in pedestrian behaviour, adapted from [11]

At the *strategic level*, pedestrians determine which activities they want to perform and the order of these activities. Given the choices at the strategic level, the *tactical level* pertains to the short-term decisions of pedestrians. Among the processes within the tactical level, route choice plays an important role and therefore a lot of literature involves this theme [12]. At last, the *operational level* is concerned mostly with the actual walking behaviour of pedestrians.

Knowledge from other disciplines such as sociology and psychology is required to understand and model the processes at the strategic and tactical level. However, as explained later in 2.3.5, evacuations procedures are well defined in time and space, as the aims and routes of pedestrians are known and usually the same. Consequently most processes at the strategic and tactical level are often considered exogenous to pedestrian simulation [11].

2.2 Walking behaviour

This section presents the empirical findings at the foundation of pedestrian traffic. The preeminent qualitative and quantitative observables are described, centred on collective phenomena and the fundamental diagram. Later an overview is presented over various modelling approaches that have been applied to the description of pedestrian dynamics.

2.3 Empirical data

2.3.1 Collective effects

Researchers of pedestrian dynamics have observed a large variety of interesting collective effects. These macroscopic effects reflect the individuals' microscopic interactions and therefore serve as a first benchmark for any modelling approach [3].

- **Jamming:** Jamming usually occurs for high densities at locations where the inflow exceeds capacity. Such locations are called *bottlenecks* and typical examples are exits or narrowings.
- **Density waves:** Density waves in pedestrian crowds are quasi-periodic density variations in space and time.
- **Lane formation:** Groups of people moving in opposite directions spontaneously form lanes of uniform walking direction. This allows for higher walking speeds and increases comfort by reducing interactions with oncoming pedestrians [13, 4].
- **Oscillations:** Oscillatory changes of the direction of motion in counterflow at bottlenecks whose frequency increases with the width and the shortness of the bottleneck [13, 4].
- **Patterns at intersections:** Alternating patterns of motion that are short lived and unstable. Some patterns, such as roundabout traffic makes pedestrian motion more efficient [4].

Figure 2.2: Jamming during a pedestrian experiment at Faculdade de Engenharia da Universidade do Porto.

Figure 2.3: Formation of lanes during a pedestrian experiment at Faculdade de Engenharia da Universidade do Porto.

- **Panic:** Various collective phenomena caused by simple reasons in emergency situations are often misleadingly attributed to panic behaviour. "Typically panic is assumed to occur in situations where people compete for scarce or dwindling resources (e.g safe space or access to an exit) which leads to selfish, asocial or even completely irrational behaviour and contagion that affects large groups. A closer investigation of many crowd disasters has revealed that most of the above characteristics have played almost no role and most of the time have not been observed at all" [3].

2.3.2 Variables that describe pedestrian traffic

The main microscopic variables that describe pedestrian dynamics are trajectories and time headways, whilst flow, density and speed are the most important macroscopic variables.

A *trajectory* is a representation of the path of a moving object over time. *Time headway* is the difference between the moments when two consecutive pedestrians cross a fixed location.

The *flow* J of a pedestrian stream refers to the number of pedestrians that cross a fixed location of per unit of time. From the various possible methods to measure flow, one from fluid dynamics is singled out: the flow through a facility of width b can be determined from the average density ρ and the average speed v of a pedestrian stream as

$$J = \rho v b = J_s b \tag{2.1}$$

where J_s is the *specific flow* and represents the flow per unit-width.

$$J_s = \rho v \tag{2.2}$$

This is the *hydrodynamic relation*, and considering traffic flow in a steady state, becomes valid for all types of flows and is known as the *fundamental relation*.

2.3.3 Fundamental diagram

The empirical relation between the macroscopic characteristics of a traffic flow is described by the *fundamental diagram*. As implied by its name, the fundamental diagram has been the focus of many investigations [14, 15, 7, 6, 8, 16, 17, 18, 19, 20, 21]. The relation quantifies the capacity of pedestrian facilities and became an elementary input for pedestrian facilities design methods [15] and used as an evaluation and benchmark of pedestrian dynamics models [22].

The fundamental diagram has three equivalent forms: $J_s(\rho), v(\rho)$ and $v(J_s)$. The shape of the diagram differs for different types of facilities like stairs and corridors, but it is assumed that for the same type of facilities, but different widths, the diagram remains the same. There are significant discrepancies between fundamental diagrams resulting from different researchers. Several explanations have been proposed such as different measurement methods [17], cultural factors [21] and differences between uni and multi-directional flow [14], but no consensus about their origin has been reached. Nevertheless, all diagrams agree that velocity decreases with increasing density.

Some special points in the diagram are:

- **Free speed** v^0 — The mean speed if $\rho = 0$ and $J = 0$.
- **Capacity** J_c — The maximum flow.
- **Capacity density** ρ_c — Density when $J = J_c$.
- **Capacity speed** v_c — Mean speed when $J = J_c$.
- **Jam density** ρ_j — Density when $v = 0$ and $J = 0$.

The fundamental diagram allows for a simple representation of the dependence of traffic characteristics on some individual pedestrian and external aspects. In his thesis, Daamen [11] summarizes some of these results from existing literature, which are presented in Figure 2.5.

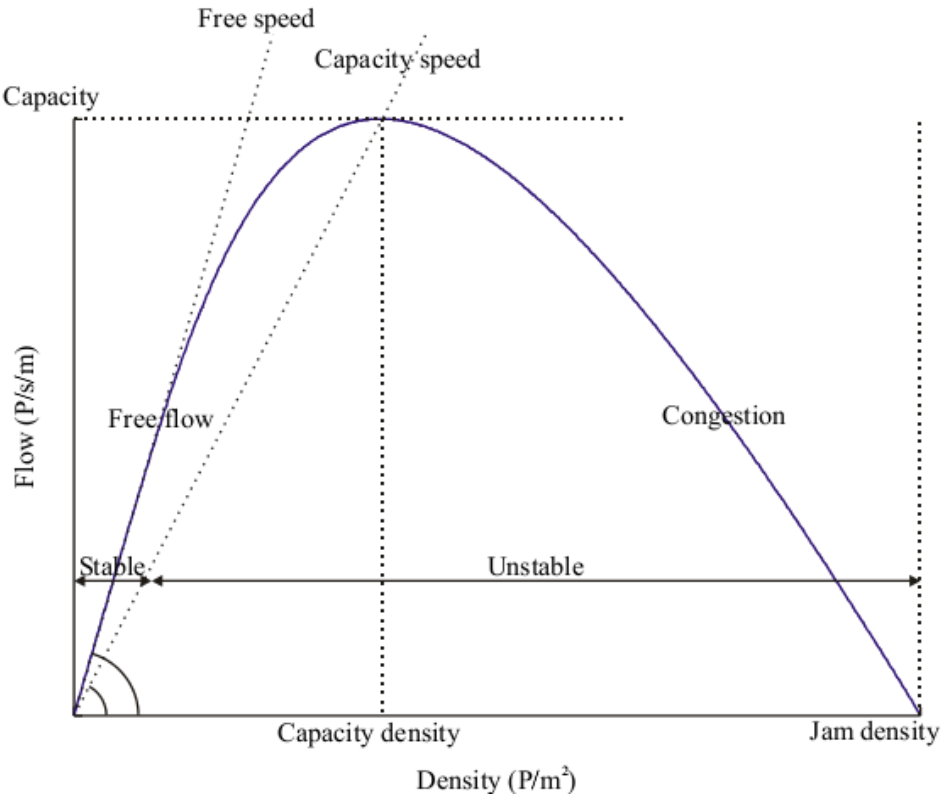


Figure 2.4: Flow-density relation for pedestrian traffic. Adapted from [11].

2.3.3.1 Fundamental diagram of pedestrian movement on a plane

Figure 2.6 presents the empirical relation between velocity and density for pedestrian movement on a plane according to Weidmann [8].

Seyfried et al. [20] discuss possible causes for the variation of slope for different density-domains:

- **Domain I: free flow** — $\rho < 0.7$ Small and increasing decline of the velocity. In this domain velocity is mostly determined by the individual free velocity of pedestrians. Passing manoeuvres are possible but cause a decrease in velocity.

- **Domain II: unstable region** — $0.7 \leq \rho < 2.3$ Velocity decreases linearly with density. Passing manoeuvres become hardly feasible, yet there is enough space to avoid contacts with other pedestrians.

- **Domain III: stable region** — $2.3 \leq \rho < 4.7$ With increasing density the velocity remains mostly constant. This might be explained by passing into marching in lock-steps and optimized usage of available space. Contacts with other pedestrians become hardly avoidable.

- **Domain IV: congestion** — $\rho \geq 4.7$ Velocity declines rapidly. Available space is strongly restricted.

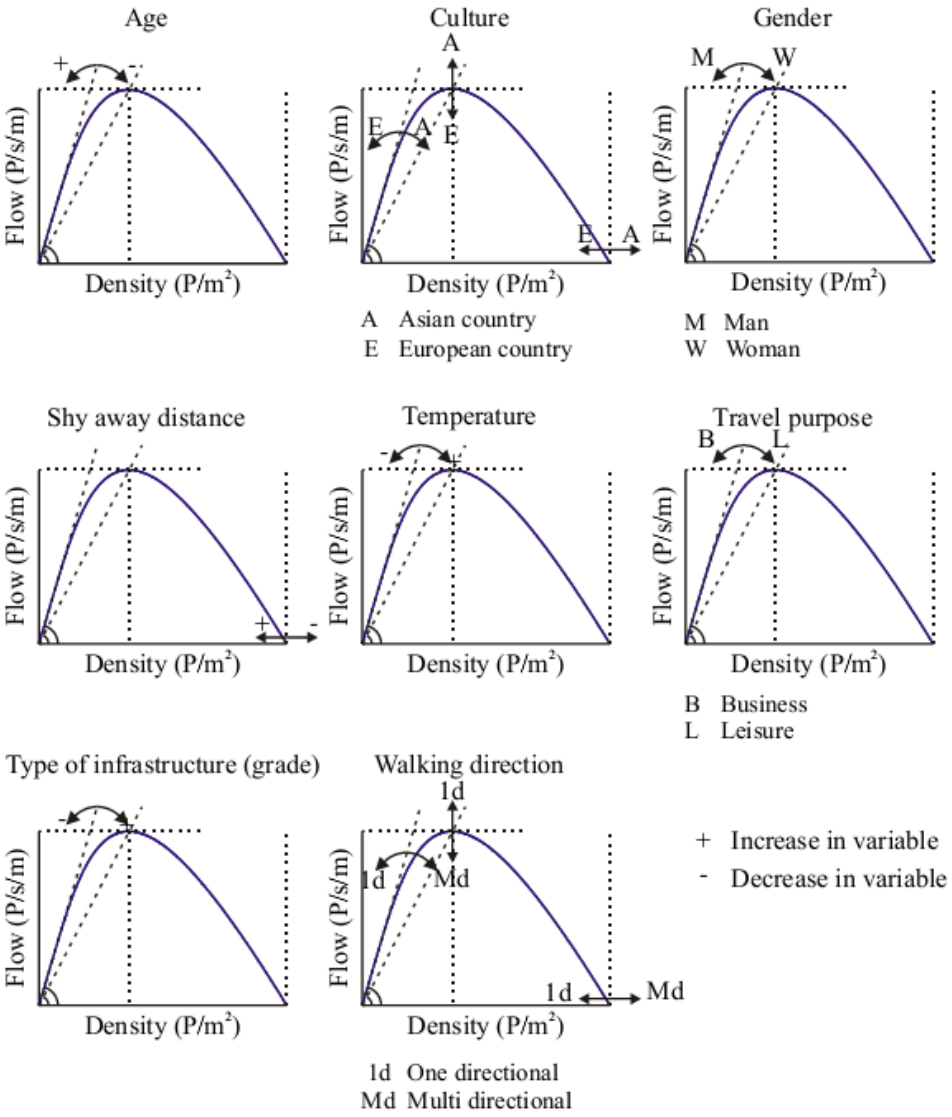


Figure 2.5: Influence of individual pedestrian characteristics and external conditions on the fundamental diagram. From [11].

2.3.3.2 Measurement methods for flow, density and velocity

The way velocities, densities or time gaps are measured present several problems in conformance to the two usual definitions of flow [3]. According to equation 2.1 the measurement of density is associated with an instantaneous mean value over space, while in other popular definitions flow is measured as a mean value over time at a certain location. Averaging these quantities over long times or large spaces reduces resolution and inhibits advantages brought by new technologies [23].

Four different measurement methods are detailed next.

Method A

In this first method, flow is measured as a mean value over a time interval Δt at a certain cross-

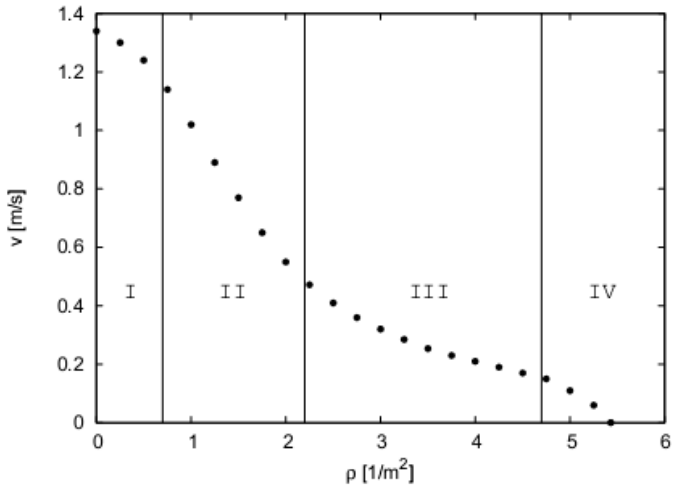


Figure 2.6: Flow-density relation for pedestrian traffic. Adapted from [20].

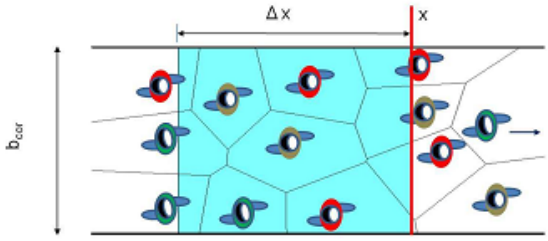


Figure 2.7: Different measurement methods. Method A is a local measurement at the cross-section with position x, while methods B-D measure the average results over space Δx. From [17].

section with position x. The velocity vi of each pedestrian passing x can be obtained directly

$$\langle J \rangle_{\Delta t} = \frac{N_{\Delta t}}{t_{N_{\Delta t}}} \quad and \quad \langle v \rangle_{\Delta t} = \frac{1}{N_{\Delta t}} \sum_{i=1}^{N_{\Delta t}} v_i(t) \tag{2.3}$$

where $N_{\Delta t}$ is the number of pedestrians crossing the location x during the interval Δt. $t_{N_{\Delta t}}$ is the elapsed time between the first and the last of the $N_{\Delta t}$ pedestrians. The time mean velocity $\langle v \rangle_{\Delta t}$ is defined as the mean value of the instantaneous velocities $v_i(t)$ of the $N_{\Delta t}$ pedestrians:

$$v_i(t) = \frac{x_i(t + \Delta t'/2) - x_i(t - \Delta t'/2)}{\Delta t'} \tag{2.4}$$

Method B

In this method velocity is measured as a mean value over space and time. A segment Δx is taken as the measurement area. The velocity $\langle v_i \rangle$ of each pedestrian is defined as the length Δx of the measurement area divided by the time needed to cross the area

$$\langle v \rangle_i = \frac{\Delta x}{t_{out} - t_{in}} \tag{2.5}$$

where t_{in} and $t_{out}t$ are the moments a pedestrian enters and exits the measurement area. The density ρ_i for each person is obtained as

$$\langle \rho \rangle_i = \frac{1}{t_{out} - t_{in}} \cdot \int_{t_{in}}^{t_{out}} \frac{N'(t)}{b_{cor} \cdot \Delta x} dt \tag{2.6}$$

b_{cor} is the width of the measurement area while $N'(t)$ is the number of pedestrians in the area at time t.

Method C

This is the classical method. Density $\langle \rho \rangle_{\Delta x}$ is defined as the number of pedestrians in the measurement section per unit of area

$$\langle \rho \rangle_{\Delta x} = \frac{N}{b_{cor} \cdot \Delta x} \tag{2.7}$$

The spatial mean velocity is the average of the instantaneous velocities $v_i(t)$ for all pedestrians in the measurement area at time t

$$\langle v \rangle_{\Delta x} = \frac{1}{N} \sum_{i=1}^{N} v_i(t) \tag{2.8}$$

Method D

This method is based on a special kind of decomposition of a metric space determined by distances to a set of objects, which form a *Voronoi diagram*. At any time, the Voronoi diagram contains a set of cells that are generated from the positions of each pedestrian. Each Voronoi cell area A_i can be thought as the personal space belonging to each pedestrian j. The density and velocity distribution over space can be defined as

$$\rho_{xy} = 1/A_i \quad and \quad v_{xy} = v_i(t) \quad if (x,y) \in A_i \tag{2.9}$$

where $v_i(t)$ is the instantaneous velocity of each person, see eq. 2.4. The Voronoi density and velocity for a specific measurement area $A_m = b_{cor} \cdot \Delta x$ is defined as

$$\langle \rho \rangle_v (x,y,t) = \frac{\int\int \rho_{xy} dxdy}{A_m} \tag{2.10}$$

$$\langle v \rangle_v (x,y,t) = \frac{\int\int v_{xy} dxdy}{A_m} \tag{2.11}$$

The specific flow

$$\langle J_s \rangle_v (x,y,t) = \langle \rho \rangle_v (x,y,t) \cdot \langle v \rangle_v (x,y,t) \tag{2.12}$$

can be computed from the Voronoi density and velocity.

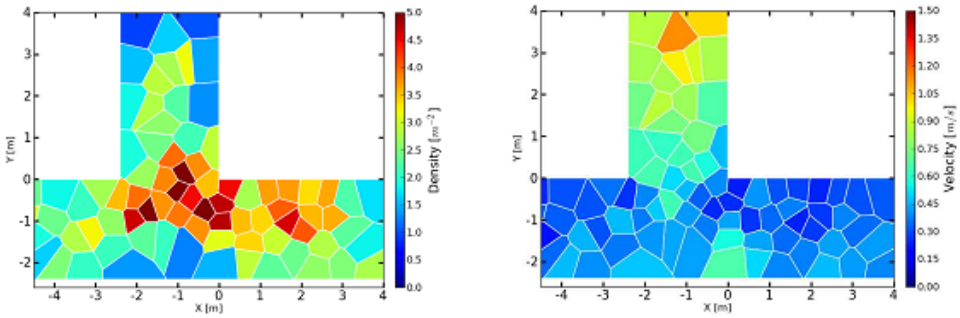


Figure 2.8: Density and velocity distribution over space obtained from Voronoi method. From [24].

2.3.3.3 Influence of the measurement method on the resulting fundamental diagram

Zhang et al. [17] concluded that the results of the fundamental diagram of pedestrian movement obtained from the same set of trajectories by these different methods agree well and that the main differences are the range of the fluctuations, which are larger in methods B and D and the resolution in time of the results, lower for method A, these results can be seen in Figure 2.9.

The *Voronoi method* (D) presents reduced fluctuation as well as good resolution in time and space, which permits examination on scales smaller that the pedestrians [23].

2.3.4 Bottleneck flow

A bottleneck designates a limited area of reduced capacity or increased demand. Pedestrian flow in these areas exhibits interesting phenomena such as the formation of lanes at the entrance [25] and clogging and arching [26, 27].

The study of pedestrian behaviour at bottlenecks is important in environments where a change in size, which might cause a change in capacity [28]. Through the estimation of maximum capacity from their fundamental diagrams, bottlenecks are paramount in the calculation of evacuation times and building capacity.

2.3.4.1 Capacity and bottleneck width

An important and very often studied question, specially for legislation purposes, is how capacity increases with rising bottleneck width.

A stepwise increase of capacity with width is expected as a consequence of the formation of more lanes since pedestrians in independent lanes, are not influenced by those on others.

However, experiments from Hoogendoorn [30, 25] show that a self-organization phenomenon known as *zipper effect* occurs at the entrance of bottlenecks. Additionally, another study [29] found that the distance of lanes and the speed in a lane increases with bottleneck width. This optimization of available space restricts the independence of lanes and leads to a very weak dependence of density and velocity inside the bottleneck. Therefore a linear increase of capacity with facility width is usually accepted [29].

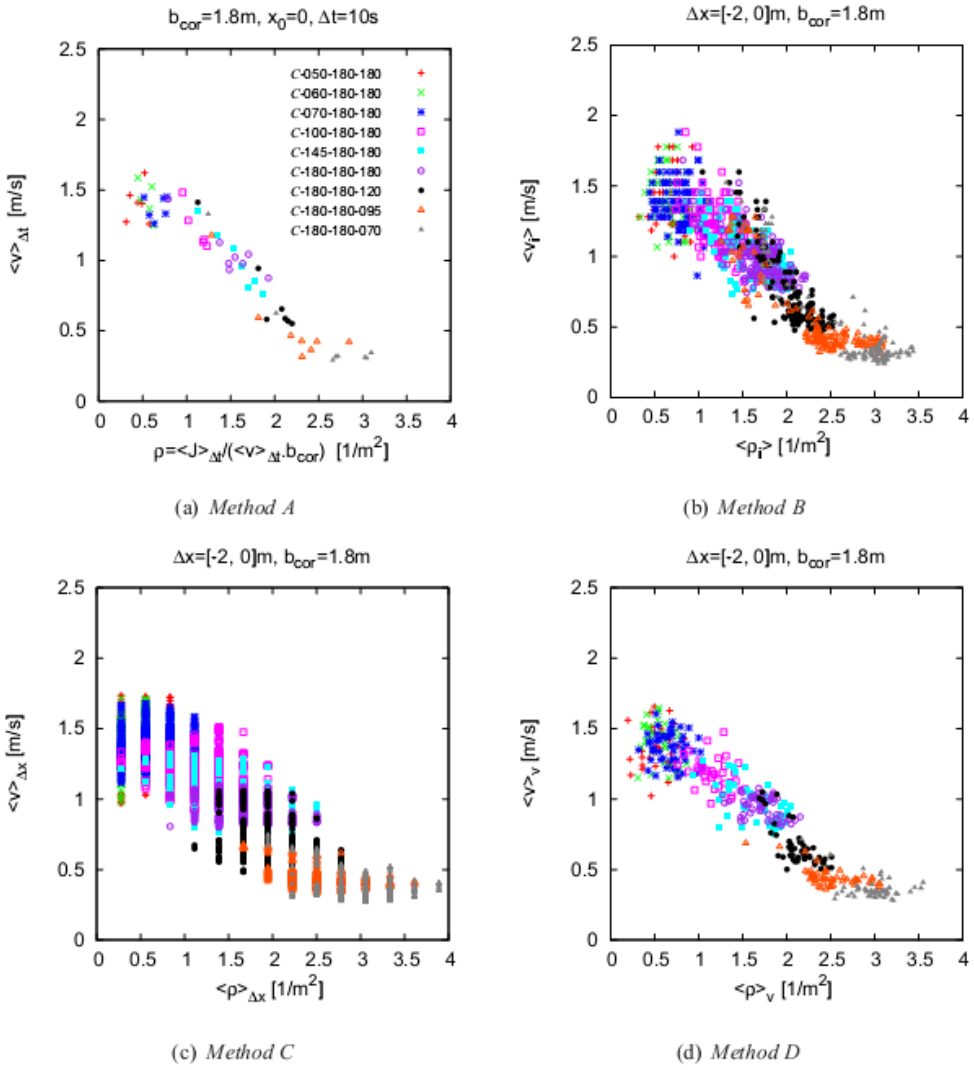


Figure 2.9: Fundamental diagrams resulting from the same set of trajectories but with different measuring methods. From [17].

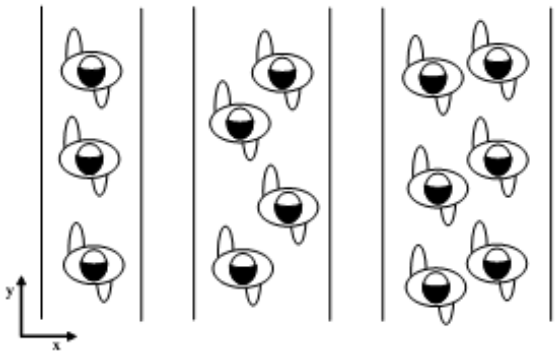


Figure 2.10: Zipper effect with increasing lane distances. From [29].

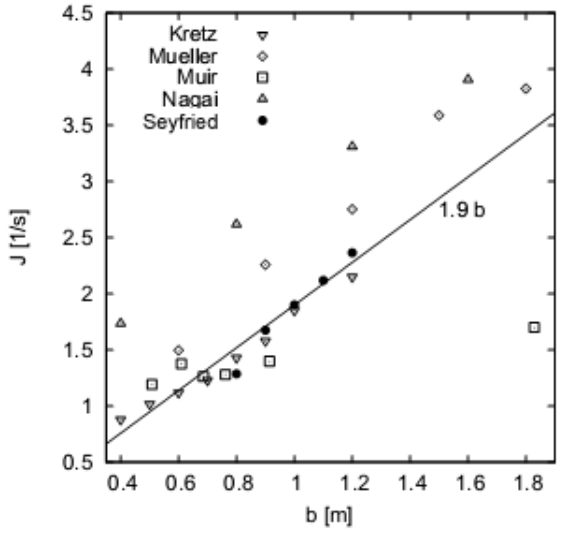


Figure 2.11: Influence of the width (b) of a bottleneck on the flow. Adapted from [29].

A comparison of several different laboratory experiments, Figure 2.11, shows that while all agree on the linearity of the width - capacity relation, data differs in the values of bottleneck capacity. The study concluded that the geometry of the bottleneck is only a minor influence on flow while different initial densities in front of the bottleneck can change the results [29].

2.3.4.2 Relation between bottleneck flow and the fundamental diagram

Another interesting question is the relation between bottleneck flow and the fundamental diagram. Notably, some of the measured values of flow in bottlenecks exceeds the maximum of empirical fundamental diagrams. This occurrence is related to the *jamming*. A jam occurs if the inflow exceeds the capacity of a bottleneck. In the case of a jam the density inside the jam will be higher that the capacity density and therefore the reduced flow in front of the bottleneck causes a smaller flow through the bottleneck than the bottleneck capacity [31].

2.3.4.3 Arching

Under some high pressure situations, where pedestrians try to pass a bottleneck as fast as possible and compete for access of the limited resource that a bottleneck represents, stable obstructions occur. This is analogous to the phenomenon of *arching* that occurs in the flow of granular materials through narrow openings [3].

Helbing et al. [33] describe a similar phenomenon, *freezing by heating*, in situations involving pedestrians under extreme conditions.

2.3.5 Evacuations

The previous empirical results presented are relative to pedestrian motion in simple scenarios. However, those results, especially bottleneck capacities, are extremely important as full-scale de-

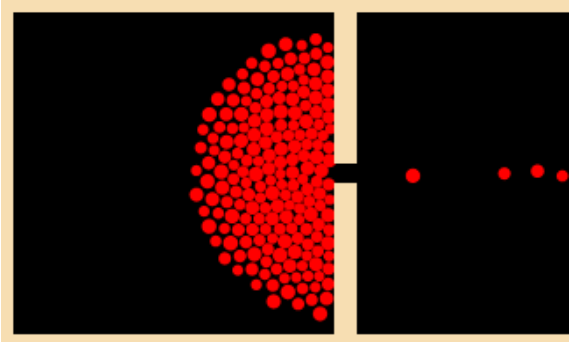


Figure 2.12: Representation of the arching phenomenon. From [32].

scriptions of evacuations from large buildings are typically a combination of many of the simple elements [3, 34].

An evacuation process is strictly limited in space and time, as the aims and routes of pedestrians are known and usually the same (exits and egress routes). Five different temporal phases can be distinguished [35]: detection time, awareness time, decision time, reaction time and movement time.

Concerning the different sources of data on evacuation processes, Schadschneider et al. [3] present a classification scheme, shown in Figure 2.13, while asserting that only data from uncontrolled or emergency situations can be used in the context of evacuation assessment.

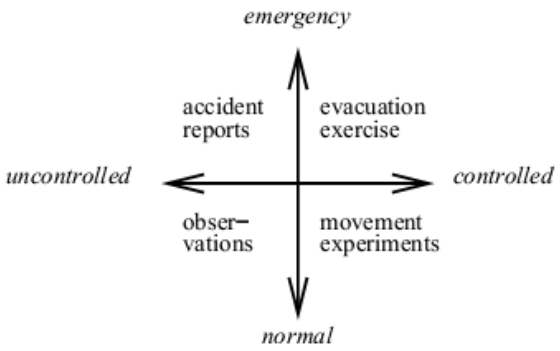


Figure 2.13: Classification of empirical data. From [3].

Data for the evacuation of complete buildings is available from two different sources: real evacuations and full scale evacuation trials. Reports of real evacuation processes are obtained from eye-witness records and incident investigations. Evacuation trials are usually observed and videotaped. These exercises are, however, often too expensive, time consuming and dangerous to be a standard measure for evacuation analysis [3].

2.3.6 Purpose of the data

In evacuation processes and modelling in general, empirical data plays a crucial role in [36]:

- **Parameter identification** — Identify factors that influence the evacuation process, e.g. bottleneck widths and capacities;
- **Calibration** — Quantify the identified parameters, e.g. flow through a bottleneck in persons per meter and second;
- **Validation** — Validate simulation results, e.g. compare the overall evacuation time measured in an evacuation with simulation or calculation results.

2.4 Modelling

2.4.1 Model characterization

Many different approaches have been developed for modelling pedestrian behaviour, which may be categorized according to the following properties [37, 38, 39]:

- **Microscopic vs. macroscopic:** Microscopic models represent each pedestrian separately while in macroscopic models the system is described through flow, density and speed relationships.
- **Discrete vs. continuous:** Models can be fully discrete, continuous or a combination of both. This means that each of the three fundamental variables that describe the system, namely space, time, and state variable can be either discrete or continuous. In continuous models pedestrian behaviour is represented by differential equations describing their movement.
- **Rule-based vs. reactive:** In the reactive approach, interactions between pedestrians are based on attraction and repulsion forces and described as equations of motion , although the forces are not necessarily physical forces. For rule-based models pedestrians make decisions based on their current situation and that in their neighbourhood, as well as their goals. There rules are often determined by arguments from psychology.
- **Deterministic vs. stochastic:** In deterministic models, past behaviour completely determines the behaviour at a certain time. In contrast, for stochastic models, the behaviour is controlled by certain probabilities such that different scenarios may occur for same situation.

The remainder of this section gives some details on existing pedestrian simulation models.

2.4.2 Fluid-dynamic and gaskinetic models

Henderson [40] proposed one of the earliest modelling approach for pedestrian dynamics, by using an analogy with fluid or gas dynamics to describe how density and velocity of pedestrian flow change overtime.

Fluid dynamic models take the macroscopic approach and are characterized by continuous variables and deterministic force based dynamics.

This modelling approach follows the similarities of pedestrian traffic flow and fluids: at low densities, pedestrians are free to move and crowd dynamics can be partially compared with the behaviour of gases. At medium or high densities however, some analogies with the motion of fluids can be made such as trails of footprints of pedestrians look similar to streamlines of fluids and emergence of pedestrian streams through standing crowds appears analogous to the formation of river beds. At high densities, the observations can be compared with granular flows.

As a result, it can be said that fluid-dynamic analogies work reasonably well in normal situations, while granular aspects dominate at extreme densities [13].

2.4.3 Social-force models

Social force models suggest that the pedestrian movement can be described as if pedestrians were subject to social forces. These forces represent internal motivations of the individual to perform certain actions [22]. These models are continuous, deterministic, and take the microscopic approach.

Social-force based models calculate the movement of individuals through the mathematical formulas associated with the application of forces, and therefore tend to result in simulations that look closer to particle animation that human movement, with agents vibrating and forgoing social rules [9].

2.4.4 Cellular automata

Cellular automata are rule-based dynamical models that are fully discrete. CA models take the microscopic approach, and its dynamics are usually rule-based and stochastic. Being rule-based has allowed complex aspects of fields such as psychology to be included in the dynamics of CA models in a simple way.

these models are based on the principle of entities (automata) occupying cells according to neighbourhood rules of occupancy, allowing only discrete movement and positioning [41]. This modelling approach has lately seen wide adoption, due to its simplicity, low computational cost and flexibility [3].

2.4.5 Agent-based models

Agent-based models are based on CA models, but try to address the limitation of homogeneous participants. The modifications to the CA approach to accommodate individual heterogeneous characteristics are so deep that the resulting models are much similar to Multi Agent Systems (MAS).

These models are characterized by the presence of autonomous entities (agents) whose action and interaction determines the evolution of the system.

2.4.6 Summary

Table 2.1 presents an summary of existing pedestrian models according to the properties from 2.4.1.

Table 2.1: Summary of pedestrian modelling

Approach	Scale	Environment	deter/stoch	Interactions
Fluid-dynamic and gaskinetic	macroscopic	continuous	deterministic	reactive
Social-forces	microscopic	continuous	deterministic	reactive
Cellular automata	microscopic	discrete	stochastic	rule-based
Agent-based	microscopic	continuous	stochastic	rule-based

2.5 Data collection

Due to the complex description of pedestrian motion sequence, in evacuation dynamics, usually only the vertical projection of the body of pedestrians is considered. As a result, data collection methods only need to record a single point in two-dimensional space to represent each pedestrian.

Data collection on crowds and pedestrians has been traditionally based on direct observation, photographs, time-lapse films [4]. Video has become an important tool as several applications for automatic extraction of pedestrian trajectories have been developed [5]. It is considerably better than direct observation, since the same scene can be studied over and over again, by human observation or using computer software with analysis algorithms. However, the usage of video recordings presents several limitations on the scenarios allowed for pedestrian experiments.

Besides video other technologies were also used to capture pedestrian behaviour, like Bluetooth and RFID [42], among others.

2.5.1 Ultra-wideband radio frequency systems for tracking pedestrians

A comparison of several radio frequency based systems for indoor human tracking [43] concludes that UWB systems are the most accurate and fault-tolerant systems that have a widespread usage in indoor localization.

According to Nekoogar [44], the usage of UWB based technology offers several advantages over narrowband communication systems, some of which Corrales, Candelas and Torres [45] claim are beneficial for indoor human tracking:

- Due to the short duration of UWB pulses, receivers are able to differentiate the original signals from the reflected and refracted ones, which make UWB based systems less sensitive to multipath fading.
- The low power UWB signals reside below the noise floor of typical narrowband receivers and enable UWB signals to share the frequency spectrum with other radio services with minimal or no interference.

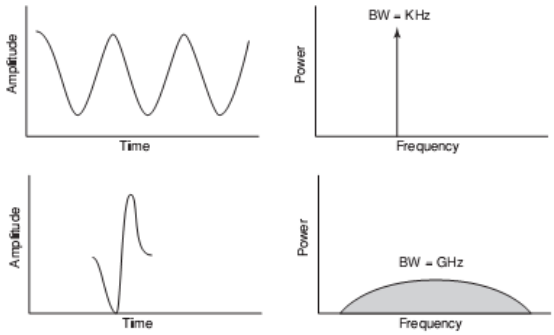


Figure 2.14: Comparison of narrowband (top) and ultra-wideband (bottom) signals in the time and frequency domains. Adapted from [44].

- There are no line-of-sight restrictions due to the long wavelength, low frequencies included in the broad range of the UWB frequency spectrum ability to penetrate a variety of materials.

- As UWB transmission is carrierless, no modulation is required, and the low-powered pulses eliminate the need for a power amplifier, resulting in simple transceiver architecture and reduced infrastructure.

The usage of a radio frequency based system for data collection allows for a number of advantages over traditional methods like automatic extraction of pedestrian trajectories from video recordings: as previously mentioned, there are no line-of-sight restrictions and its also suited for low ceiling buildings. It also encompasses a wider breath of possible simulation scenarios, like limited visibility situations such as dark or smoke filled rooms, which are very common in emergency situations due to fire.

Assigning identifiable tags to individual participants, also allows to easily associate some characteristics of the individual pedestrian (e.g. gender, height and age) with its trajectory. This might allow better understanding of the dynamics of heterogeneous crowds and study the effect of outliers like elderly people or people with mobility impairments

2.6 Knowledge Discovery in Databases

The process of knowledge discovery in databases (KDD) is the answer for the analysis of the immense data volume in scientific and operational databases. While traditional analytic techniques falter, KDD attempts to turn raw data into information, allowing the usage of that information to gather knowledge about the specific domain. KDD involves sweeping databases using some sort of algorithm or heuristic as guide and usually consists of multiple, connected steps, including data selection, data preprocessing, incorporation of prior knowledge, data mining, visual representation, interpretation and evaluation of the results [46].

2.6.1 Datamining

The primary goals of data mining are prediction and description. Prediction is associated with using some known variables to predict unknown or future values [46]. Description focuses on identifying valid, potentially useful, and understandable correlations and patterns in existing data [47].

To achieve the stated goals, data mining can involve tasks from the following classes: deviation, association, clustering, classification, regression and generalization.

- **Association** comprises of discovering interesting and significant relations and dependencies between variables;

- **Classification** can be defined as the task of assigning objects to one of several predefined categories, by learning a target function f that maps each attribute set x to one of the predefined class labels y [48];

- **Clustering** divides data into groups (clusters) that are meaningful, useful, or both. The clusters should capture the natural structure of the data [48];

- **Deviation** involves discovering the most significant deviations of the data from the expected or previously measured values;

- **Generalization** consists of finding a compact description for the data;

- **Regression** encloses mapping a data item to a real-valued prediction variable [46];

Table 2.2: Summary of Data-Mining Tasks and Techniques. Adapted from [49]

Knowledge type	**Description**	**Techniques**
Classification	Predict the class label that a set of data belongs to based on some training datasets	Bayesian classification Decision tree induction Artificial neural networks Support vector machines
Clustering	Determine a finite set of implicit groups that describe the data	Cluster analysis
Association	Find relationships among item-sets or association / correlation rules, or predict the value of some attribute based on the value of other attributes	Association rules Bayesian networks
Deviation	Find data items that exhibit unusual deviations from expectations	Clustering Outlier detections Evolution analysis
Regression	Lines and curves summarizing the database	Regression Sequential pattern extraction
Generalization	Compact description of the data	Summary rules Attribute-oriented induction

2.6.2 Spatio-temporal data mining

Spatio-temporal data mining can be defined as the extraction of implicit knowledge, spatial and temporal relationships, or other patterns not explicitly stored in spatio-temporal databases [50].

The data that populates spatio-temporal databases might come from robotics and computer vision applications, GIS, CAD, biology and mobile computing as well as temporal data obtained by registering events and monitoring processes.

The need to investigate both the temporal and spatial dimensions add complexity to the data mining tasks, as spatial relations (distance, direction, shape, topology) as well as temporal relations must be considered in the data mining methods. Some of these relations are not explicitly defined in the dataset and must be extracted from the data, either before the mining process starts or on-the-fly as they are needed.

Extraction of spatial and temporal relations introduces fuzziness that might have influence on the results, that can also be impacted by the spatial resolution and temporal granularity of the data. Other challenge is concerned with data representations, as working at the level of the stored data (points, lines, timestamps) is often undesirable, and complex transformations are required to describe them at higher conceptual levels.

Chapter 3

The Methodological Approach

3.1 Problem Statement

As stated in the previous chapters, this project aims gather data from the movement of pedestrians and subsequently extract information to discover knowledge such as patterns, trends, dependencies and rules of movement. This information can then be used as input for tools to help predict the behaviour of masses and design buildings and plan emergency egress.

Although Agent-based models of pedestrian dynamics are pivotal in predicting the dynamical properties of large human crowds, these are often built upon analogies with the dynamics of fluids or movement of particles in a gas and therefore ignore psychological factors that guide pedestrian behaviour. Some models try to overcome this limitation with the usage forces between particles that arise from empirical rules and insights, but ultimately lack information that might be extracted from the analysis of real pedestrian movement. There is also a clear necessity for reliable data against which the models can be calibrated and to ensure their validity.

Data collection on crowds and pedestrians has been traditionally based on direct observation and time-lapse films. Video has become an important tool as several applications for automatic extraction of pedestrian trajectories have been developed. However, the usage of video recordings presents several limitations on the scenarios allowed for pedestrian experiments, some of which, in particular line of sight and visibility restrictions, can be overcome radio frequency based data collection systems. The availability of one such system (*Ubisense*) in the LIACC made its use compelling for this task. It is noteworthy to mention that there is little literature regarding tracking humans trajectories in confined spaces and dealing with the problems that might arise such as the precision of the extracted trajectory.

The optimal source of data for evacuation processes is real evacuations. However data collected from real evacuations comes mostly from witness records and incident investigations as the unpredictability of such events does not allow for the installation of radio frequency or video based tracking systems that would also require each person to wear specific markers or devices. As such, there is no reliable way to gather evacuees trajectories. Evacuation exercises, overcome the data collection problem, but are often too expensive, time-consuming and dangerous.

The first main problem then is to gather reliable data on pedestrian movement with the usage of the *Ubisense* real time location system as gathering data from evacuations is difficult. After this, the data must be analysed and the problem resides in extracting meaningful information and discovering knowledge for later usage in model validation and calibration for the 'mSPEED' framework.

3.2 Overview

This section focuses on how to tackle and try to solve the above stated problem.

Regarding the need for data on pedestrian movement to model human behaviour, and the associated challenges that arise from gathering data from real evacuations and conducting full-scale evacuation exercises, we chose to devise and perform small scale walking behaviour experiments with volunteers. These experiments are conducted under controlled conditions and take place in different scenarios, aiming to replicate several situations found in real facilities such as bottlenecks, corners and junctions. As described during the literature review, more complex scenarios are often a combinations of many of the simple elements such as these. This approach deals with the reduced number of participants available, space related limitations, and the danger and expensiveness of the time consuming full scale evacuation exercises. Another advantage of performing experiments is the possibility to vary each influencing variable to examine its effects on the behaviour of pedestrians.

The next challenge is related with recording the movement of pedestrians in these experiments. Towards this, we propose the usage of an ultra-wideband based indoor location system, *Ubisense*, that allows gathering spatio-temporal data from each participant in each of the performed experiments in the form of trajectories. The location system must then be studied and understood to ensure that the resulting data is of the highest quality possible. Important topics to cover in this step include the system installation, sensor and tag placement, as well as using an application to interface with the *Ubisense* API to extract and record the location information of each pedestrian.

Once the experiments are over, arises the problem of dealing with and trying to improve the quality of the data, so that analysis can be performed. Although the location system used can, under ideal conditions, achieve an accuracy of up to 20 cm, previous experiments and test scenarios could only reach sub-meter accuracy [51]. To deal with this problem we will reject corrupted readings and filter out data bellow a certain quality threshold, while also trying to reduce the noise in the readings. Consequently, an application will be developed to help visualize the extracted data by reconstructing the experiments in a three-dimensional simulation scenario.

The next step is concerned with extracting meaningful data from the gathered trajectories. Towards this goal we will use spatio-temporal analysis and visualization to make sense of the data. Analysis will be performed by calculating basic motion descriptors such as velocities, acceleration, measures of sinuosity, path length and evacuation time to characterize trajectories. Visualization will focus on the representation of trajectories as well as the distribution of some of the quantified properties along each trajectory or the experiment area.

Concerning the final goal of behavioural explanation, we face the challenges of how to extract patterns from the trajectory data and how to explain behaviour with the extracted patterns. To solve this challenges we plan to partition trajectories into segments or sub-trajectories that share the similar properties. This step will be performed taking as input the characterization obtained from the previous step, by using motion descriptors. Finally we plan to aggregate the sub-trajectories into clusters that symbolize a certain behaviour. The tasks that make up the solution for this last goal - infer human behaviour from pedestrian trajectories - will be made possible with the application of spatio-temporal data mining and machine learning concepts.

3.2.1 The Ubisense Real Time Location System

The *Ubisense real-time location system* is an in-building ultra-wideband radio based tracking system which can obtain accurate information of the positions of people and objects. This system uses small devices (tags, Figure 3.2b) that send UWB pulses to a network of hardware receivers (sensors, Figure 3.2a) fixated in the localization area.

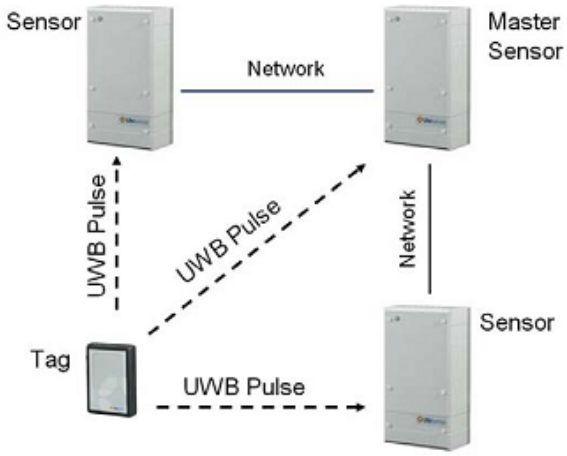


Figure 3.1: Ubisense components. From [52].

The system applies a combination of TDOA (Time-Difference of Arrival) and AOA (Angle of Arrival) techniques to estimate the position of each tag [52]. These tags are small devices that can be attached to objects or carried by personnel. Sensors can also be connected to a computer, and Ubisense also provides a middleware platform which can manage and filter real-time location information and simplify the creation of location aware applications that monitor several localization areas simultaneously, as schematized in Figure 3.3.

Section 2.5.1 already mentioned some of the advantages of using a UWB based technology for indoor human tracking, comparing it to both other radio-frequency based technologies as well as the conventional video based method used in most experiments of pedestrian movement. The system availability in the LIACC is another reason for its usage as the tool for data collection in this project.

(a) Sensor

(b) Tag

Figure 3.2: Ubisense hardware components in detail

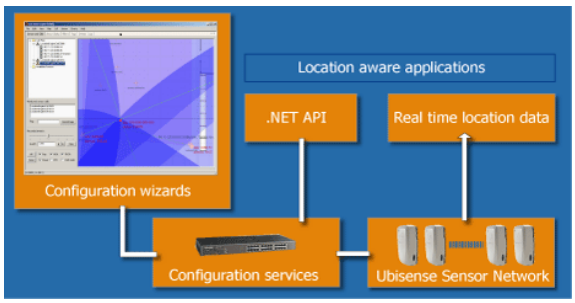


Figure 3.3: Ubisense location engine

The Ubisense location system has seen wide adoption in the industry, especially in manufacturing plants, providing location services for tracking assets in order to improve and better control processes [53]. Lately, it has also been used to track personnel during military and fire fighter training and operations, and as a behaviour analysis tool based on coordinates of body tags [54].

3.2.2 Spatio-temporal Analysis and Visual Inspection

Towards understanding how movement patterns relate to the characteristics of the facilities and crowds, it is fundamental to make sense of the data. A simple and effective way to do this is through visualization, as the human brain is used to perform similar analysis on a daily basis. This can be achieved by using spatio-temporal analysis and visualization techniques that represent trajectories with associated data such as velocities and accelerations.

The techniques to be explored rely on the description of trajectories and mobile objects through the calculation of different values and indices, as well as their visual representation. These descriptors provide essential movement characteristics of mobile objects and therefore are an important step to compare the movement of different pedestrians and therefore differentiate behaviour.

We will focus on some quantitative and qualitative representations of mobile objects; quantitative consist on quantification by basic motion descriptors such as velocities, accelerations, orientation. Some aggregated indexes will also be calculated such as measures of tortuosity and path length, evacuation time and average velocity. Qualitative representation consists on the visualization of trajectories associated with the quantitative data as well as data inferred for the facility, to better understand spatio-temporal pattern of the movement of pedestrians.

3.2.3 Spatio-temporal Data Mining

A important task left is to analyse and understand the spatio-temporal patterns, processes, and behaviours of pedestrians, in order to extract useful and meaningful information and knowledge about pedestrian dynamics.

The information provided by the previous task, spatio-temporal analysis, is an initial step towards this goal, but knowledge discovery techniques can still be applied to find hidden patterns, relations and behaviours in the data. The approach to contextualize movement patterns follows. Its aim is to extract hidden patterns, trends, behavioural contexts and useful information and knowledge from the trajectory data collected beforehand. Trajectory data mining discovers spatio-temporal knowledge through exercises including pattern detection, clustering, classification, generalization, outlier detection and visualization.

A important input for this step are the descriptions of trajectories obtained in the previous step and they act as features for the data. The trajectory data mining scheme employs trajectory partitioning and clustering algorithms to extract behavioural patterns of pedestrians using multiple descriptors.

The main challenges consist of describing and characterizing trajectories in order to be able to extract patterns as well as explaining behaviour based on the extracted patterns. As mentioned before, the first phase is dealt in the previous chapter, Spatio-temporal Analysis and Visualization, where pedestrian trajectories are described by calculation basic motion descriptors, which provide movement characteristics of the pedestrians.

Towards the second challenge, as humans can present multiple movement behaviours across space and time, to describe behaviour from trajectory data it is better to capture local motion behaviours rather than to use aggregated motion. Trajectory partitioning is then used to decompose a trajectory into a set of sub-trajectories with similar motion characteristics. The resulting sub-trajectories can consequently be clustered into groups that share the same local behaviours.

3.3 Methodology

3.3.1 Pedestrian Experiments

Our aim for the experiments described in this section is twofold. First, we aim to provide valid data for pedestrian dynamics model elicitation, as well as model validation in different facilities Another goal is to evaluate the usage of a UWB based real-time location system for pedestrian movement data collection and trajectory extraction.

Towards meeting these goals, we devised a set of experiments in simple scenarios that mimic sections that compose more complex facilities. The simple scenarios chosen were: corridor, bottleneck, corner, and t-junction. Different configurations such as different corridor widths or varying number of participants were also explored.

These experiments follow a set of similar experiments performed in Germany at the Jülich Supercomputing Centre [20, 29, 16, 24, 55], in which pedestrian trajectories were accurately extracted from video recordings. The similarity between our experiments and others found in literature allows comparison, which is valuable to evaluate the novel method for trajectory extraction.

The choice of simple scenarios is also related with some limitations faced when planning the experiments: only 30 *Ubisense* tags are available, imposing a limit on the maximum number of participants. There are also space limitations: as the system and equipment installation and setup is a lengthy process, it is not possible to perform the experiments outdoor, and a room with full availability for an entire week is required, as well as the needed materials to build the barriers that limit the experiment area.

The experiments were divided in three general set-ups: single-file, narrow passage and corner and T-junction. Each set-up aims to provide data suitable to study different phenomena: fundamental relation in a simple scenario, unidirectional pedestrian flow through bottlenecks and more complex configurations like corners and merging of flows respectively.

3.3.1.1 The Single-File Scenario

The reduced degrees of freedom associated with movement along a line (single-file movement) helps reduce the number of effects that might influence the relation between density and velocity of pedestrian movement. Single-file movement is therefore the simplest system for investigation of this dependency, which is the goal of the first experiment. To avoid boundary effects and limit the amount of participants needed, this first scenario is composed of a looping track with corridor width such that it does not impede the free movement of arms but enforces single-file movement by preventing passing and the formation of multiple lanes (Figure 3.4).

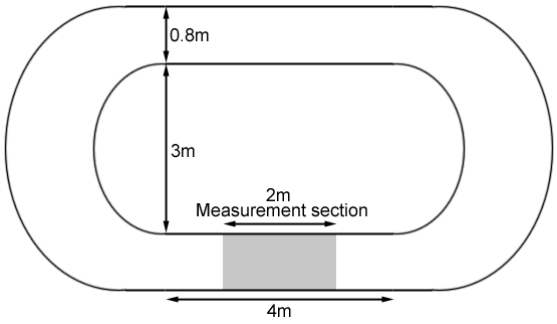


Figure 3.4: Experimental setup for the single file scenario

Test subjects were instructed not to overtake and were distributed uniformly in the corridor. Each experiment run had each person complete two full loops before leaving the track. To regulate pedestrian density, three runs with 10, 15 and 20 randomly chosen pedestrians were performed. Ten runs with a single pedestrian were also performed with the purpose of free velocity determination.

3.3.1.2 Bottleneck Scenario

In the second scenario, focus is given to unidirectional pedestrian flow through bottlenecks (Figure 3.5). We plan to study flow as a parameter of bottleneck width as well as the formation of arc like structures as density at the front of the bottleneck increases. With these experiments we aim to extrapolate information about capacity estimation in pedestrian facilities, as this is an important parameter that allows the estimation of time taken in an evacuation scenario, or minimum width in bottlenecks for different desired evacuation durations.

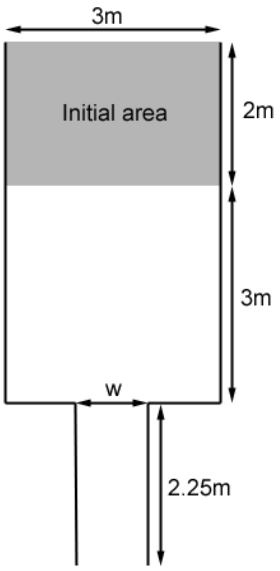
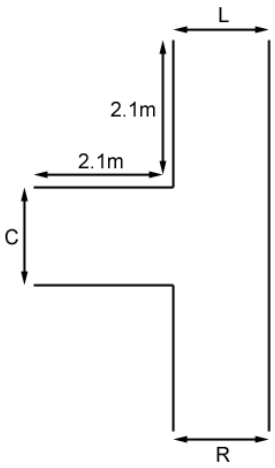


Figure 3.5: Experimental setup for the bottleneck scenario

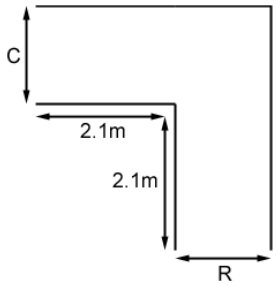
Before each runs starts, participants are placed in the starting area to ensure a constant initial density between different experiments runs. Ensuring constant initial density is important as it influences flow through the bottleneck [29]. After the signal to start the experiment, participants move through a bottleneck with sufficient height to assure constant width from the hips to the shoulders of the test persons. Four runs with different bottleneck widths $w = 0.8, 0.9, 1.0$ and $1.4m$ and 20 participants were performed.

3.3.1.3 Corner and T-Junction Scenarios

While reliable data is crucial for modeling and calibration of pedestrian dynamics models, it is almost nonexistent for complex types of facilities like T-junctions or corners [24]. The last set of experiments aim to combat this deficiency by improving the data base related to pedestrian dynamics and investigate the bottleneck flow due to merging, splitting and turning of streams. For this last scenario, different experiments were carried out with different corridor widths behind and in front of merging and splitting of streams (Figure 3.6). In each of the different corridor sections C, L or R, width is set either to 0.9 or $1.5m$. The corner angle is always kept at $90°$.



(a) T-junction



(b) Corner

Figure 3.6: Experimental setup for the corner and T-junction scenarios

In the corner experiments, pedestrians start behind the entrance to the corridor and move across in a single direction. Different runs are recorded with varying corridor widths behind and in front of the corner.

For some T-junction experiments, a similar number of pedestrians moves from both left and right simultaneously, merging into the main stream, while in other experiments the initial stream composed of all pedestrians is split as participants walk into different corridors.

3.3.2 Data collection

In contrast to the data collection method performed in most pedestrian experiments, trajectory extraction from video recordings, on our experiments automatic data collection is performed with the aid of the above mentioned UWB tracking system, by assigning individual tags to participants, whose position can then be then tracked.

The experiments are to be conducted with up to 30 participants and performed in the Faculty of Engineering of the University of Porto. The experiments should be confined to an open area with as few metallic elements as possible, to minimize electromagnetic interference and ensure good readings from the RF system.

3.3.2.1 Tracking system installation and configuration

As stated in 3.2.1, the chosen tracking system contains both hardware and software parts. The hardware part is composed of active tags that emit ultra-wideband (UWB) pulses and receivers which are used to calculate the location of the tags based on the received UWB signal. Several sensors are chained together to form a cell, within which tags can be localized. Each sensor determines the Angle of Arrival (AOA) of the incoming signal and the Time Difference Of Arrival (TDOA) is obtained with a timing cable between the sensors. Within each cell one sensor is

responsible to collect and process the information from the other sensors and generate location events which are fed to the *Location Engine* over a network.

The *Location Engine* is a software service responsible to process location the data and pass information to other applications via an API. In conjunction with other software components such as the *Location Engine Configuration*, *Map*, *Platform Control*, *Security Manager*, *Service Installer*, *Service Manager* and *Site manager*, they make up the *Ubisense Location Platform* software suite that allows the set-up and calibration of location sensors and tags, as well as their configuration into cells and the defining the location system's behaviour.

For the physical set-up of the tracking system, after being secured at a height of $2.3m$ from the ground, four sensors were placed in the four vertices of the bounding rectangle containing the area where experiments took place (Figure 3.7). Within this area, the different scenarios were built using tables and a vinyl foldable wall.

Figure 3.7: Ubisense sensors placement in the experiment area

Tag placement plays an important role in ensuring that good readings will be achievable. Tags become difficult to read when they are in close proximity to materials that absorb a large amount of radio frequency energy, such as water, that makes up most of the human body. For the purposes of the experiments, tags were attached on top of Christmas hats secured with straps (Figure 3.8).

Location data about the placement of the sensors was then input into the *Location Engine Configuration* tool to create a *cell* within which the location of *tags* can be obtained. This tool also features an automatic calibration procedure where the data about sensors orientation and offset is calculated from location events of *tags* in known positions within the *cell*. To improve the location data, activity thresholds, which are the incident power levels for UWB signals above which tag sightings are assumed valid rather than background noise, can be set for each sensor, and filters to reject outlier location events and provide robust location tracking can be set for each tag.

The filters are algorithms for estimating tag positions from sensor measurements which maintain a state that is used to reject outlier measurements and reduce noise. Several parameters control the behaviour of the algorithm such as constraints on the motion of tags.

The filter algorithms available in the *Location Engine Configuration* are variants of *Information filtering*, which uses the previous motion of the tag to predict its position at the time a new

Figure 3.8: Ubisense tags attached to hats

sighting is made. The sensor data is then compared to the prediction and a new estimate for the position of the tag is created. Therefore the algorithm requires a model for the motion of tags, and four variants are provided, each with different constraints on the motion model, as summarized in table 3.1.

Table 3.1: Stateful filters

Name	Filter state dimensions	Dynamics of the tracked tag
Static fixed height information filtering	2 dimensions (2D position)	Mostly still or moving unpredictably. Height fixed to small range.
Static information filtering	3 dimensions (3D position)	Mostly still or moving unpredictably. Position unconstrained.
Fixed height information filtering	4 dimensions (2D position and velocity)	The tag is moving with a predictable speed. Height fixed to a small range.
Information filtering	6 dimensions (3D position and velocity)	The tag is moving with a predictable speed. Position entirely unconstrained.

Adapted from [56]

None of the scenarios for pedestrian experiments features elevation changes, therefore the tags attached to the pedestrians will keep a constant distance to the floor and tag movement can be considered to be only horizontal. As recommended by the system instruction manual, the algorithm that provides the most constrained motion allowed by the tracked object was chosen - *fixed height information filtering*. With this filter, the tag is free to move horizontally, but vertical motion is constrained to be close to a fixed height above the cell floor. If a tag is not seen for some

period of time, it is assumed to continue moving with the same horizontal speed and direction, with uncertainty increasing over time, but to keep close to the same fixed height.

3.3.2.2 Data extraction and recording

While the location system captures events that provide information about the spatial position of tags, a special application has to be developed to record these events so that trajectories may be stored in a file or database. This application was already developed before this project, although some modifications were be made to attempt to extract more information about the positioning events. Its interface is presented in Figure 3.9

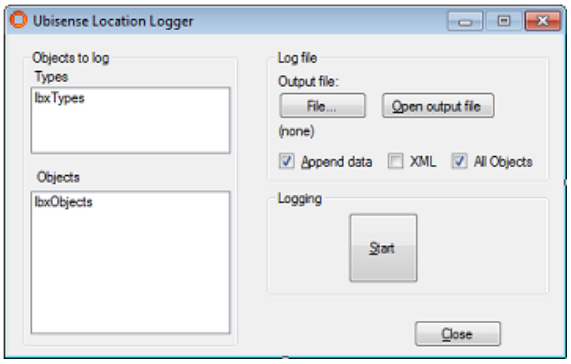


Figure 3.9: Ubisense location logger.

3.3.3 Visualization tool

A simple visualization tool was developed to offer a first insight into the gathered data. The tool takes the files with location data provided by the previously mentioned data recording tool as input and provides a three-dimensional representation of the position pedestrians at any time during each experiment as can be seen in Figure 3.10. It also features the ability to playback the experiments at different speeds and show only the desired tags.

Although it presents limited use for analysis, as only collective phenomena can be observed, this visualization tool is useful in the initial stage of the project to understand possible shortcomings or problems with the data early on.

3.3.4 Data Filtering and Cleansing

Despite the filtering algorithms provided by the tracking system, visualization of early samples of trajectory data have shown that the collected trajectories were still noisy and imprecise and not yet adequate for the desired movement analysis. The samples also contained some useless information such as unused tags outside the measurement area that were tracked during experiments and the data recorded in the instants just before the start and after the end of the experiment duration.

To improve the quality of the recorded trajectories a series of actions are performed on the data:

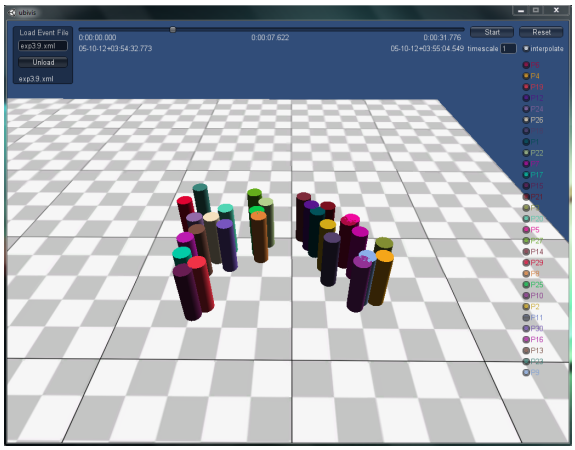


Figure 3.10: Visualization tool.

- **Cleansing:** With the aid of the previously proposed visualization tool, unused and malfunctioning tags in each data record file can be found and discarded. The precise instants for the start and end of each experiment can also be defined, to better trim trajectory data;

- **Filtering:** The data can contain some corrupted readings that can be rejected through filtering;

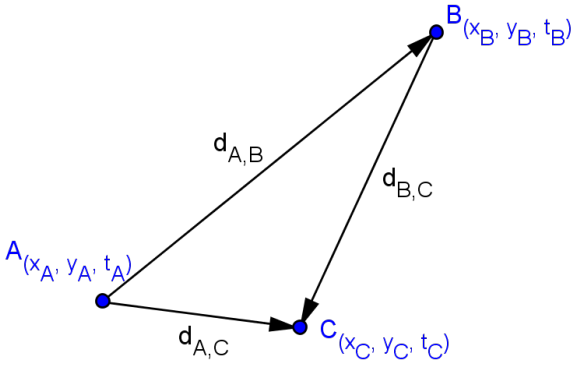
- **Smoothing:** As information about the whole trajectory is available, smoothing techniques can be used to reduce the noise in the readings by improving the location of each measurement using information about previous and subsequent sightings;

- **Re-sampling:** The collected data is sampled at irregular time intervals, and fixed interval re-sampling is required for some of the analysis techniques later presented;

3.3.4.1 Data filtering

The filtering technique employed is based on some modifications to the Maximum Redundant Distance filter (MRD) of the Douglas Argos-filter algorithm [57]. The MRD filter works by retaining locations based on spatial redundancy between consecutive locations. As the distances between tracked locations are small in the present work, a similar approach was developed using velocity information instead of spatial distance.

Each step of the algorithm considers three consecutive locations A, B and C as presented in Figure 3.11. If any of the velocities of the movement between positions AB, BC or AC are lower that a certain maximum velocity max_v, then the endpoints of the movement are marked as retained. The algorithm continues by considering the locations B, C and D in the next step and so on until the last location. At the end all locations not marked as retained are removed from the trajectory.



$$\text{if } \frac{|d_{A,B}|}{t_B - t_A} \leq max_v \text{ then retain A and B}$$
$$\text{if } \frac{|d_{B,C}|}{t_C - t_B} \leq max_v \text{ then retain B and C}$$
$$\text{if } \frac{|d_{A,C}|}{t_C - t_A} \leq max_v \text{ then retain A and C}$$

Figure 3.11: Filter logic

3.3.4.2 Data smoothing

Smoothing is applied to reduce and smooth-out short-term irregularity in the data series to reveal more clearly the underlying trend in the data. For this purpose, smoothing was achieved trough the usage of a weighted moving average smoother.

The weighed moving average smoother uses information from a sample window around each point in the data to obtain a better estimation of the point. It has multiplying factors to give different weights to data at different positions in the sample window. The weighted moving average of window size m estimates locations $\hat{y}(i)$ as:

$$\hat{y}(i) = \sum_{j=-k}^{k} a_{i,j} y_{i+j} \tag{3.1}$$

Where $k = (m-1)/2$ and the weights $[a_{i,-k}, ..., a_{i,k}]$ sum to one and depend on the temporal distance between the point in consideration y_i and the weighted point y_{i+j}:

$$a_{i,j} = \frac{|t_k - t_{i-k}|}{t_{i+k} - t_{i-k}} \tag{3.2}$$

3.3.5 Initial analysis

3.3.5.1 Analysis of the single file experiment

The main goal of the single file experiment was to investigate the dependency between the density and velocity of pedestrian movement.

Towards this goal, only a straight section with length $l_m = 2m$ (Fig. 3.4) is considered to determine the density-velocity relation of pedestrian movement. Entrance (t^{en}) and exit (t^{ex}) times are recorded for each pedestrian crossing the entrance (x^{en}) and exit (x^{ex}) of this section. From these times, both the average velocity v_i (3.3) of each crossing i as well as the number of persons inside the measurement section $N(t)$ at each instant t can be obtained.

$$v_i = \frac{l_m}{t_i^{ex} - t_i^{en}} \tag{3.3}$$

Taking into consideration the large period between consecutive measurements for each tag, a linear interpolation between the positions and instants when each tag is first located inside (x_i^{in}, t_i^{in}) or outside (x_i^{out}, t_i^{out}) the measurement section and the positions and times associated with the previous locations $(x_i^{in-1}, t_i^{in-1}, x_i^{out-1}, t_i^{out-1})$ are necessary to better estimate the exact entrance (3.4) and exit (3.5) times.

$$t_i^{en} = t_i^{in-1} + \left[\left(t_i^{in} - t_i^{in-1} \right) \frac{x^{en} - x_i^{in-1}}{x_i^{in} - x_i^{in-1}} \right] \tag{3.4}$$

$$t_i^{ex} = t_i^{out-1} + \left[\left(t_i^{out} - t_i^{out-1} \right) \frac{x^{ex} - x_i^{out-1}}{x_i^{out} - x_i^{out-1}} \right] \tag{3.5}$$

Density at each instant t can be obtained from the instantaneous number of pedestrians N inside the measurement section:

$$\rho(t) = N(t)/l_m$$

As the measurement section is short, only small numbers of persons can be inside. Consequently the value of density calculated from the above definition jumps between discrete values. An enhanced definition of the density, calculated through the time headways between successive pedestrians avoids this problem, but its calculation is a challenging task as the data collection system is unable to provide the location of two tags at the same instant, and also because inaccuracies would increase as error from two different measurements would have to be taken into account.

The density assigned to each pedestrian crossing the measurement section is determined as the mean value of density during the crossing:

$$\rho_i = \frac{1}{t_i^{ex} - t_i^{en}} \int_{t_i^{en}}^{t_i^{ex}} \rho_n dt$$

The differences between the mean value of density over time determined by this method or by the time headway method is negligible [21].

The fundamental diagram of pedestrian movement is finally obtained by taking the graphical representation of the velocity (v_i) - density (ρ_i) pairs of each crossing. An example of such representation is in Figure 3.12, which was obtained by Seyfried et al. [20] using the same method.

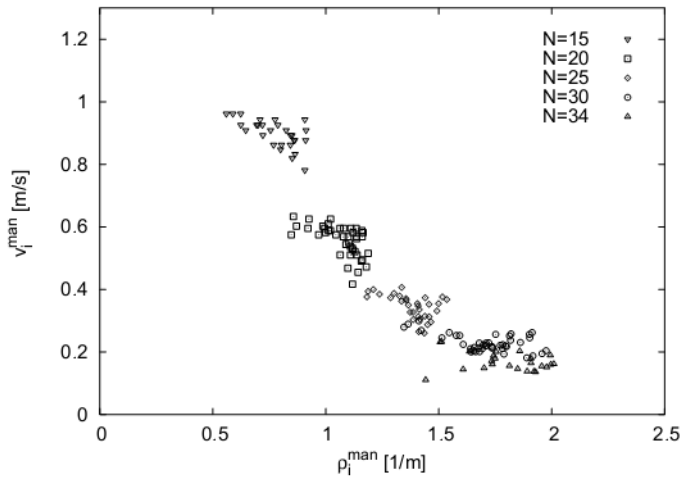


Figure 3.12: Representation of the velocity (v_i) - density (ρ_i) pairs of each crossing for runs with various number of participants (N). From [20].

3.3.6 Spatio-temporal Analysis and Visual Inspection

This part of the project attempts to use a set of spatial and spatio-temporal analysis methods to measure and compare movement. The analysis will be explore the concepts of velocity and acceleration, sinuosity, directional statistics and density. The first step involves the calculation of a variety of values that describe the participants movement in an aggregate manner, such as average velocity, total displacement and evacuation time for each pedestrian.

In order to contextualize the trajectories, motion descriptors such as instantaneous velocity and orientation are calculated for each tag at different instants for the entire duration of the experiment. This is an important step towards describing trajectories in ways that allow a deeper analysis and comparison.

Finally the data gathered during these steps is represented graphically so empirical and visual review can be performed.

3.3.6.1 Trajectory representation

A trajectory is a continuous function of time which, given a time instant t, returns the position at time t of the object in a d-dimensional space. In our application, however, only a finite set of observations of the moving object exist. Moreover, the location system used provides location events at irregular rates and without temporal alignment between the sightings of different objects. For an approximate reconstruction of the trajectory, tags are assumed have a piecewise linear movement: moving along a straight line with constant speed between observations.

The data collected from each experiment is a collection of trajectories, $\tau = \{TR_1, ..., TR_n\}$, where each trajectory TR_i represents the movement of each of the n pedestrians carrying a UWB tag. A trajectory can then be stored as sequence of multidimensional points $TR_i = p_1, p_2, ..., p_m$, where m is the number of points in the trajectory. As the movement is constrained to a horizontal plane, only two spatial dimensions and another for temporal data are needed for each point $p_j = x_j, y_j, t_j$.

3.3.6.2 Velocity and acceleration

The general properties of movement retlative to a fixed point or prior speed are described by the *velocity* and *acceleration*. These properties of motion can differentiate distinct behaviours of moving objects; for example, velocity can explain states of mobile objects such as walk, stay, run and drive.

For an object's two dimensional vector moving from point P to point Q, the displacement is the change in the position vector r, given the x and y component of Δr as Δx and Δy, and Δt referring to the duration of the described motion.

$$\begin{aligned} \Delta x = &\quad x_Q - x_P \\ \Delta y = &\quad y_Q - y_P \\ \Delta t = &\quad t_Q - t_P \end{aligned}$$

The average velocity v_{av} is the vector quantity equal to the displacement divided by the time interval.

$$v_{av} = \frac{\Delta r}{\Delta t}$$

The average acceleration, v_{av} of an object from point P to point Q is the vector change in velocity, Δv, divided by elapsed time.

$$a_{av} = \frac{\Delta v}{\Delta t}$$

3.3.6.3 Tortuosity

Tortuosity is a commonly used characteristic to describe movement paths which consists in how much tortuous and twisted a path is in a given space or time. Studies on the tortuosity of trajectories have been done mainly in the fields of biology to research the movement of animals. Throughout the years, a variety of tortuosity indices have been proposed such as Mean Squared Distance, Fractal D, Straightness Index and Sinuosity [58]. The *Straightness Index*, promoted by Batschelet [59], is a particularly interesting measure due to its simplicity and intuitiveness: it consists on the ratio of the distance between the beginning and end points of the path (D) and the total path length (L).

$$SI = \frac{D}{L} \tag{3.6}$$

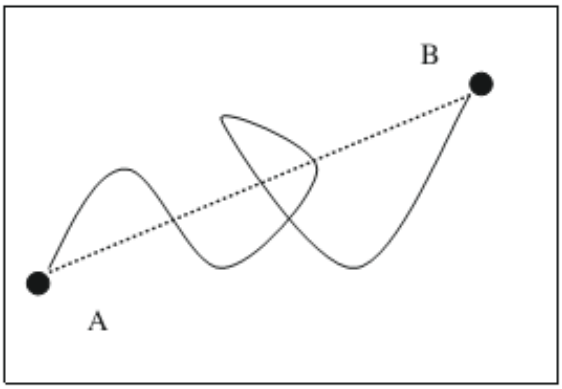


Figure 3.13: Straightness index. Point A represents the start of a trajectory and B the end of it. The dotted straight line is the displacement (distance between the beginning and end points of the path) while the sinuous line represents the recorded trajectory. From [60].

The Straightness index ranges from 0 to 1, where higher values mean straighter paths.

3.3.6.4 Directional statistics

Directional data is an important part of the information contained in trajectories. Techniques from linear statistics, however, are not appropriate in characterizing a sample of angular observations. Even a technique as basic as computing a mean may lead to disturbing results [61].

Measures such as the directional mean and circular variance are analogous to familiar measures from linear statistics and help examine the directional pattern on trajectories. The directional mean ($\bar{\theta}$) is a measure of central tendency calculated as follows:

Let $\theta_1, \theta_2, ..., \theta_n$ be a sample on n directions

$$\bar{\theta} = \begin{cases} tan^{-1}\left(\frac{\bar{S}}{\bar{C}}\right), & \text{if } \bar{S} > 0 \text{ and } \bar{C} > 0 \\ tan^{-1}\left(\frac{\bar{S}}{\bar{C}}\right) + \pi, & \text{if } \bar{S} > 0 \text{ and } \bar{C} < 0 \\ tan^{-1}\left(\frac{\bar{S}}{\bar{C}}\right) + 2\pi, & \text{if } \bar{S} < 0 \text{ and } \bar{C} > 0 \end{cases} \tag{3.7}$$

Where $\bar{S}$ and $\bar{C}$ are the mean of the x and y coordinates:

$$\bar{S} = \frac{1}{n}\left(\sum_{i=1}^{n} \sin\theta_i\right)$$
$$\bar{C} = \frac{1}{n}\left(\sum_{i=1}^{n} \cos\theta_i\right)$$

The circular variance (S_v) indicates the variability of the directions of the movement and can be obtained from the *mean resultant length* ($\bar{R}$) as follows:

$$\bar{R} = \sqrt{\bar{S}^2 + \bar{C}^2}$$
$$S_v = 1 - \frac{\bar{R}}{n} \tag{3.8}$$

Unlike its linear analogue, the values for circular variance are always between 0 and 1; values close to 0 mean that all vectors point in the same direction while values near 1 indicate a wide range of directions.

3.3.6.5 Density

The surrounding environment of pedestrians conditions their movement and therefore it is vital to try to gather and incorporate this type of external information when describing pedestrian movement, in order to understand the resulting behaviour. An important part of each pedestrian surroundings is how space around him is occupied - how many other pedestrians are nearby and how close they are. Density presents a simple way to incorporate this data, however different ways to calculate density have been explored, some of them detailed in 2.3.3.2.

From the presented methods, the one involving the Voronoi decomposition of space was chosen for the purpose of density calculation. In this method, partitions of space are determined by distances to a set of objects, which form a *Voronoi diagram*. At any time, the Voronoi diagram contains a set of cells that are generated from the positions of each pedestrian. Each Voronoi cell area A_i can be thought as the personal space belonging to each pedestrian j as illustrated in Figure 3.14. The density over space can be defined as:

$$\rho_{xy} = 1/A_i \qquad if(x,y) \in A_i \tag{3.9}$$

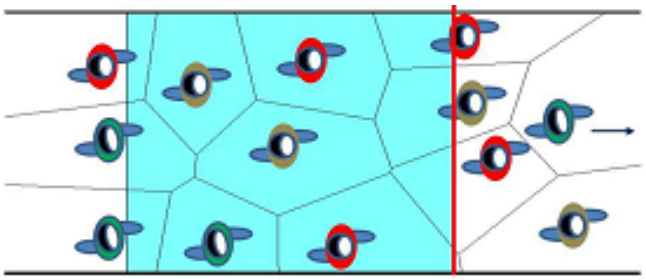


Figure 3.14: Voronoi partition of space. Adapted from [24].

3.3.6.6 Aggregate trajectory characterization

In each of the experiments, for each of the participating pedestrian's trajectories, the following values that hold information about that whole trajectory were obtained: displacement, evacuation time, trajectory length, average velocity, straightness index, directional mean and circular variance. Table 3.2 shows how these values were calculated.

Table 3.2: Aggregate motion descriptors of trajectories

Value	Formula
Displacement	$D = \sqrt{(x_m - x_1)^2 + (y_m - y_1)^2}$
Evacuation Time	$ET = t_m - t_1$
Path Length	$L = \sum_{j=2}^{m-1} \sqrt{(x_{j+1} - x_j)^2 + (y_{j+1} - y_j)^2}$
Average Velocity	$V_{av} = L/ET$
Straightness Index	3.6
Directional Mean	3.7
Circular Variance	3.8

3.3.6.7 Segmented trajectory characterization

Given the assumed representation for trajectories as a set of points with an associated location and time in 3.3.6.1, each trajectory can be seen as composed of $(m-1)$ segments, where each segment s_k is connects two sequential points (ps_{k1}, ps_{k2}).

Other information can be computed and associated to each new segment of the trajectory in order to further improve the description of movement. Simple values to obtain are velocities and accelerations, as well as the orientation. A measure of density can also be obtained through the reciprocal of the areas assigned to each cell obtained by generating a Voronoi diagram in which the location of pedestrians act as seeds 3.3.6.5. These values are stored in multi-dimensional vectors that characterize the segments of each trajectory.

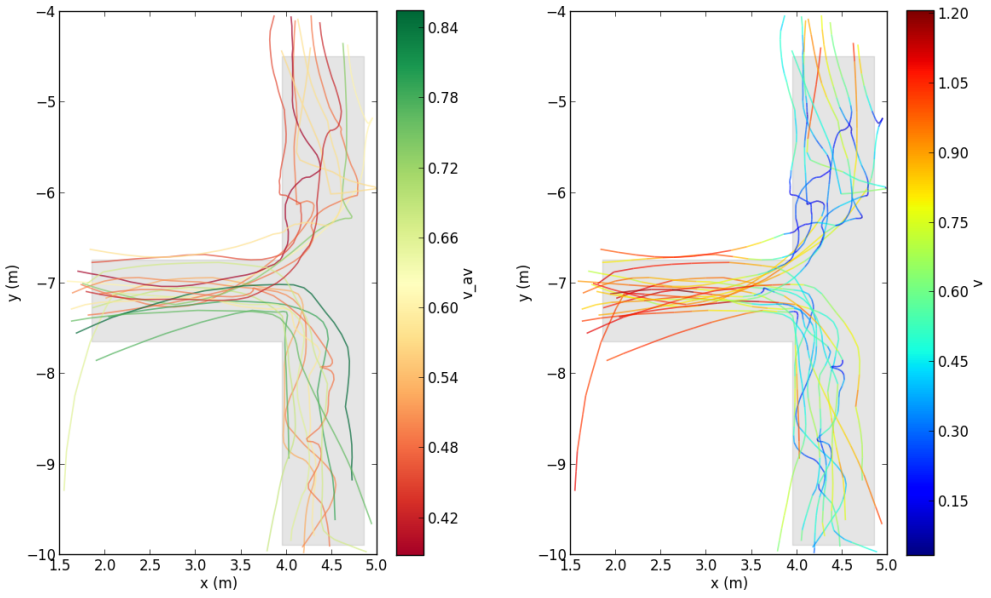
Table 3.3: Motion descriptors of trajectory segments

Value	Formula
Displacement	$D_k = \sqrt{(x_{k2} - x_{k1})^2 + (y_{k2} - y_{k1})^2}$
Velocity	$V_k = D_k / t_k$
Acceleration	$(V_{k+1} - V_k)/t_k$
Orientation	$\theta_k = arctan2(y_{k2} - y_{k1}, x_{k2} - x_{k1})$
Density	3.9

3.3.6.8 Visualization

In order to comprehend the relations between movement patterns and the characteristics of the facilities and behaviour of crowds it is essential to extract meaning from the data. A practical technique to help do so is visualization, since the human brain is so visually oriented and performs analogous analysis on a day to day basis.

In order to accomplish this, the data computed during from the previously mentioned spatio-temporal analysis can be presented through visual representations, where, for instance, the data can be associated with the individual pedestrian by drawing the pedestrian's trajectory in a certain colour, that corresponds to a value for the associated characteristic. Figure 3.15a shows an example of such representation. In a similar fashion, other representations show different characteristics for each segment that composes the trajectories, resulting in figures such as Figure 3.15b.



(a) Trajectories colour coded by average velocity

(b) Trajectories colour coded by velocity in each segment

Figure 3.15: Examples of the visualization of trajectories

Towards obtaining information for the whole experiment area, kernel density estimation can be

used. Kernel density estimation (KDE) is an approach to achieve a non-parametric estimation of data density [62]. Given a set of n data samples x_i, the kernel density estimator $\hat{f}_h(x)$ is calculated as:

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i)$$

based on a kernel function K and a bandwidth parameter h. Appropriate selection of K where $\int K(x)dx = 1$ and $K(x) \geq 0$ allows interpretation of $\hat{f}_h(x)$ as a density function that approximates the probability distribution function $f(x)$ of the data items x_i from which it has been constructed. Bandwidth h is a parameter with influences the smoothness of the density reconstruction.

The extension of KDE into two dimensions leads to express the estimate of probability density at the point (x, y) as:

$$\hat{f_{2D}}(x,y) = \frac{1}{nh_1^2}\sum_{i=1}^{n} K_1\left(\frac{x - x_i}{h_1}, \frac{y - y_i}{h_1}\right) \tag{3.10}$$

Where K_1 is kernel function defined over 2-dimensional space. Examples of commonly used kernel functions are the uniform, triangular, gaussian, and Epanechnikov functions [63]. The Epanechnikov kernel function is defined in 3.11 and was the one chosen for our project.

$$K_1(x,y) = \begin{cases} \frac{2}{\pi}\left(1 - (x^2 + y^2)\right)\alpha, & \text{if } (x^2 + y^2) < 1 \\ 0, & \text{if } (x^2 + y^2) \geq 1 \end{cases} \tag{3.11}$$

In addition, the point density value can be magnified by a scaling factor (α), which can be a scalar value of motion descriptors.

3.3.7 Spatio-temporal Data Mining

The goal of this part of the project is to identify local behaviours from the trajectory dataset. Trajectories may have a long and complicated path, hence although two trajectories can be identical in some limited sections, they might not be similar as a whole. Instead of clustering trajectories as a whole, they can be segmented at the points where the behaviour of the trajectory changes rapidly - *characteristic points*. Each partition is then represented by a line segment between two consecutive characteristic points. To discover similar portions of movement, clustering is applied on the resulting set of segments by grouping sub-trajectories based on density.

This approach to trajectory clustering was introduced in [64], where the algorithm for trajectory clustering - TRACLUS is presented.

Finally, the resulting sub-trajectory clusters are characterized by employing the techniques introduced in the previous section to produce aggregate information about the sub-trajectories in each cluster.

The proposed methodological approach is presented in Figure 3.16 and involves three steps; trajectory partitioning, segment clustering and cluster classification.

1. **Trajectory partitioning**
   Trajectories are partitioned into line segments using the minimum description length (MDL) principle.

2. **Segment clustering**
   Discover sub-trajectories that indicate share common movement patterns by using a density-based clustering algorithm.

3. **Cluster characterization**
   Extract aggregate motion descriptors for sub-trajectories in the same cluster following the procedure presented in 3.3.6.6.

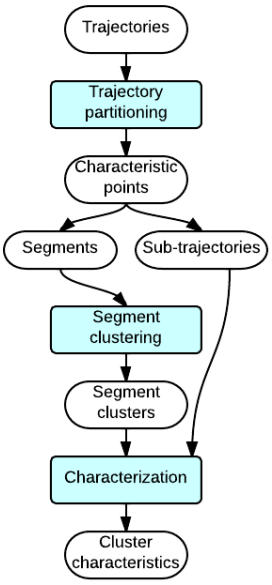


Figure 3.16: Procedure for spatio-temporal data mining

3.3.7.1 Distance function between line segments

Some of the steps of the trajectory partition and clustering algorithms require a way to measure distance between line segments. The distance function used in clustering line segments by the TRACLUS approach is composed of three components: the *perpendicular distance* ($d_\perp$), the *parallel distance* ($d_\parallel$), and the *angle distance* (d_θ). These components are illustrated in Figure 3.17.

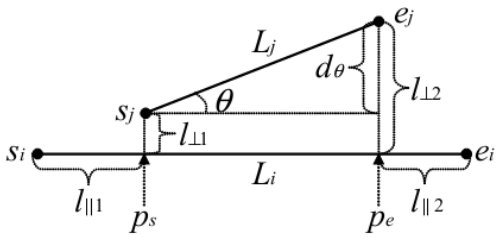


Figure 3.17: Components of the distance function for line segments. From [64].

Suppose p_s and p_e are the projections of of the points s_j and e_j onto L_i. The Euclidean distance between s_j and p_s and between e_j and p_e are $l_{\perp 1}$, $l_{\perp 2}$ respectively. The *perpendicular distance* between L_i and L_j is defined by the Lehmer Mean of $l_{\perp 1}$ and $l_{\perp 2}$ of order 2 as follows:

$$d_{\perp}(L_i, L_j) = \frac{l_{\perp 1}^2 + l_{\perp 2}^2}{l_{\perp 1} + l_{\perp 2}} \tag{3.12}$$

The minimum of the Euclidean distances of p_s to s_i and e_i is $l_{\parallel 1}$. Likewise, $l_{\parallel 2}$ is the minimum of the Euclidean distances of p_e to s_i and e_i. The *parallel distance* between L_i and L_j is defined as the minimum of the Euclidean distances of $l_{\parallel 1}$ and $l_{\parallel 2}$:

$$d_{\parallel}(L_i, L_j) = min(l_{\parallel 1}, l_{\parallel 2}) \tag{3.13}$$

The *angle distance* between L_i and L_j is defined in Formula 3.14, where $\|L_j\|$ is the length of L_j and θ is the smaller intersecting angle between L_i and L_j.

$$d_{\theta}(L_i, L_j) = \begin{cases} \|L_j\| \times sin(\theta), & \text{if } 0^\circ \leq \theta < 90^\circ \\ \|L_j\|, & \text{if } 90^\circ \leq \theta \leq 180^\circ \end{cases} \tag{3.14}$$

Finally, the distance between two line segments is defined as the weighted sum of the three distances:

$$d(L_i, L_j) = \omega_{\perp} \cdot d_{\perp}(L_i, L_j) + \omega_{\parallel} \cdot d_{\parallel}(L_i, L_j) + \omega_{\theta} \cdot d_{\theta}(L_i, L_j) \tag{3.15}$$

Where the weights $\omega_{\perp}$, $\omega_{\parallel}$ and ω_{θ} are set to 1 as default but can be changed depending on the application.

3.3.7.2 Trajectory partitioning

Trajectory partitioning aims to find a set of points at which the behaviour of a trajectory $TR_i = p_1, p_2, ..., p_m$ changes rapidly - *characteristic points* ($\{p_{c_1}, p_{c_2}, ..., p_{c_{par}}\}$). The trajectory is then segmented at each *characteristic point* into $(par-1)$ trajectory partitions, and each partition is represented by a line segment between two consecutive *characteristic points* $\{p_{c_1}p_{c_2}, p_{c_2}p_{c_3}, ..., p_{c_{par-1}}p_{c_{par}}\}$. An example of a trajectory and its partitions is illustrated in Figure 3.18

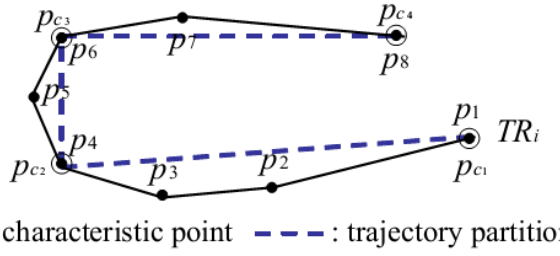


Figure 3.18: Example of a trajectory and its trajectory partitions. From [64].

The partitioning process is achieved by finding the optimal trade-off between two contradictory desirable properties: *preciseness* and *conciseness*. Preciseness alludes to the minimization of the difference between a trajectory and a set of its trajectory partitions, whereas conciseness invokes the smallest possible number of trajectory partitions. A possible method for finding the optimal tradeoff between preciseness and conciseness is based on the minimum description length (MDL) principle.

The MDL cost consists of two components: $L(H)$ and $L(D|H)$. $L(H)$ is the length, in bits, of the description of the hypothesis H, and $L(D|H)$ is the length, in bits, of the description of the data D when encoded with the help of the hypothesis. The best hypothesis to explain the data is the one that minimizes the sum of $L(H)$ and $L(D|H)$.

In the trajectory partition problem, a trajectory corresponds to D and set of trajectory partitions corresponds to H. The lengths of the hypothesis and the data are defined as:

$$L(H) = \sum_{j=1}^{par-1} log_2(len(p_{c_j}, p_{c_{j+1}})) \tag{3.16}$$

$$L(D|H) = \sum_{j=1}^{par-1} \sum_{k=c_j}^{c_{j+1}-1} \{log_2(d_\perp(p_{c_j}p_{c_{j+1}}, p_k p_{k+1})) + log_2(d_\theta(p_{c_j}p_{c_{j+1}}, p_k p_{k+1}))\} \tag{3.17}$$

Where $L(H)$ measures the conciseness and is obtained by the sum of the length of all trajectory partitions of a trajectory. $L(D|H)$ indicates the preciseness and is computed by the sum of the distances between a segment of a trajectory partition ($p_{c_j}p_{c_{j+1}}$) and each line segment ($p_k p_{k+1}$) residing in the trajectory partition.

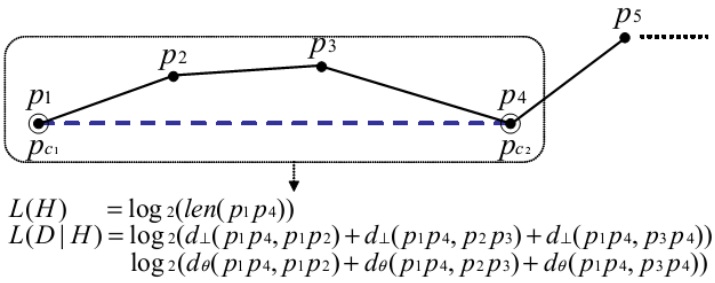


Figure 3.19: Formulation of the MDL cost. From [64].

The formulation of the MDL cost is presented in Figure 3.19. Consequently, finding the optimal trajectory partitioning is finding the hypothesis that minimizes $L(H) + L(D|H)$. However, the obligation to consider every subset of the points in a trajectory makes finding the optimal partitioning prohibitive.

In order to approximate the solution, Lee, et al. [64] defined two MDL costs, $MDL_{par}(p_i, p_j)$ and $MDL_{nopar}(p_i, p_j)$. $MDL_{par}(p_i, p_j)$ is the MDL cost of a trajectory between p_i and p_j assuming that p_i and p_j are the only characteristic points. $MDL_{nopar}(p_i, p_j)$ is the MDL cost assuming that there is no characteristic point between p_i and p_j (i.e. preserving the original trajectory).

The approximate solution is then longest trajectory partition $p_i p_j$ that satisfies $MDL_{par}(p_i, p_j) \leq MDL_{nopar}(p_i, p_j)$ for every k such that $i < k \leq j$. The approximate algorithm for partitioning a trajectory is presented next.

Algorithm 1 Approximate Trajectory Partition. Adapted from [64].

Input: A trajectory $TR_i = p_1, p_2, p_3, \ldots, p_j \ldots, p_{len_i}$
Output: A set CP_i of characteristic points

```
Add p_1 into the set CP_i                                   ▷ the starting point
startIndex ← 1, length ← 1
while startIndex + length ≤ len_i do
    currIndex ← startIndex + length
    cost_par ← MDL_par(p_startIndex, p_currIndex)
    cost_nopar ← MDL_nopar(p_startIndex, p_currIndex)
    ▷ check if partitioning at the current point makes the MDL cost larger than not partitioning
    if cost_par > cost_nopar then                           ▷ partition at the previous point
        Add p_currIndex−1 into the set CP_i
        startIndex ← currIndex − 1, length ← 1
    else
        length ← length + 1
Add p_len_i into the set CP_i                               ▷ the ending point
```

3.3.7.3 Segment clustering

The clustering algorithm for line segments employed in this work was introduced in [64] and is based on DBSCAN [65]. Let $\mathcal{D}$ denote the set of all line segments. Lee, et al. [64] summarize the notions required for density-based clustering in the following definitions:

Definition 1. The *ε-neighborhood* $N_\varepsilon(L_i)$ of a line segment $L_i \in \mathcal{D}$ is defined by $N_\varepsilon(L_i) = \{L_j \in \mathcal{D} \mid dist(L_i, L_j) \leq \varepsilon\}$.

Definition 2. A line segment $L_i \in \mathcal{D}$ is called a *core line segment* with regard to ε and *MinLns* if $|N_\varepsilon(L_i)| \geq MinLns$.

Definition 3. A line segment $L_i \in \mathcal{D}$ is *directly density-reachable* from a line segment $L_j \in \mathcal{D}$ w.r.t. ε and *MinLns* if $L_i \in N_\varepsilon(L_i)$ and $|N_\varepsilon(L_i)| \geq MinLns$.

Definition 4. A line segment $L_i \in \mathcal{D}$ is *density-reachable* from a line segment $L_j \in \mathcal{D}$ w.r.t. ε and *MinLns* if there is a chain of line segments $L_j, L_{j-1}, ..., L_{i+1}, L_i \in \mathcal{D}$ such that L_k is directly density-reachable from L_{k+1}.

Definition 5. A line segment $L_i \in \mathcal{D}$ is *density-connected* to a line segment $L_j \in \mathcal{D}$ w.r.t. ε and *MinLns* if there is a line segment L_k such that both L_i and L_j are density-reachable from L_k.

Definition 6. A non-empty subset $\mathcal{C} \subseteq \mathcal{D}$ is called a *density-connected set* if $\mathcal{C}$ satisfies the following two conditions:
(1) *Connectivity:* $\forall L_i, L_j \in \mathcal{C}$, L_i is density-connected to L_j;
(2) *Maximality:* $\forall L_i, L_j \in \mathcal{C}$, if $L_i \in \mathcal{C}$ and L_j is density-reachable from L_i, then $L_j \in \mathcal{C}$.

Definition 7. The *set of participating trajectories* of a cluster C_i is defined by $PTR(C_i) = \{TR(L_j) \mid \forall L_j \in C_i\}$. $TR(L_j)$ denotes the trajectory from witch L_j has been extracted and $|PTR(C_i)|$ is called the *trajectory cardinality* of the cluster C_i.

Given a set $\mathcal{D}$ of line segments and two parameters ε and $MinLns$, the clustering algorithm in [64] generates a set of $\mathcal{O}$ clusters. In the algorithm a cluster is defined as a density-connected set of line segments, however, only sets of line segments that present a trajectory cardinality above a certain threshold are considered as clusters, in order to ensure that they explain the behaviour of a sufficient number of trajectories. The clustering algorithm for line segments is detailed in Algorithm 2.

Algorithm 2 Line Segment Clustering. Adapted from [64].

Input: (1) A set of line segments $\mathcal{D} = \{L_1, \ldots, L_{num_{ln}}\}$
(2) Two parameters ε and $MinLns$
Output: A set of clusters $\mathcal{O} = \{C_1, \ldots, C_{num_{clus}}\}$

1: $clusterId \leftarrow 0$ ▷ an initial id
2: Mark all the line segments in $\mathcal{D}$ as *unclassified*
3: **for each** $L \in \mathcal{D}$ **do**
4: **if** L is *unclassified* **then**
5: Compute $N_\varepsilon(L)$ ▷ compute the $\varepsilon - neighborhood$ of each unclassified line segment L
6: **if** $|N_\varepsilon(L)| \geq MinLns$ **then** ▷ if L is determined as a core segment
7: Assign $clusterId$ to $\forall X \in N_\varepsilon(L)$
8: Insert $N_\varepsilon(L) - \{L\}$ into the queue $\mathcal{Q}$
9: EXPANDCLUSTER($\mathcal{Q}, clusterId, \varepsilon, MinLns$)
10: $clusterId \leftarrow clusterId + 1$ ▷ a new id
11: **else**
12: Mark L as *noise*
13: Allocate $\forall L \in \mathcal{D}$ to its cluster $\mathcal{C}_{clusterId}$
14: **for each** $\mathcal{C} \in \mathcal{O}$ **do** ▷ check the trajectory cardinality of each cluster
15: **if** $|PTR(\mathcal{C})| < MinLns$ **then** ▷ a threshold other than MinLns can be used
16: Remove $\mathcal{C}$ from the set $\mathcal{O}$ of clusters

17: **procedure** EXPANDCLUSTER($\mathcal{Q}, clusterId, \varepsilon, MinLns$) ▷ compute a density-connected set
18: **while** $\mathcal{Q} \neq \varnothing$ **do**
19: Let M be the first line segment in $\mathcal{Q}$
20: Compute $N_\varepsilon(M)$
21: **if** $|N_\varepsilon(L)| \geq MinLns$ **then**
22: **for each** $X \in N_\varepsilon(M)$ **do**
23: **if** X is *unclassified* or *noise* **then**
24: Assign $clusterId$ to X
25: **if** X is *unclassified* **then**
26: Insert X into the queue $\mathcal{Q}$
27: Remove M from the queue $\mathcal{Q}$

Chapter 4

Experimental Results, Analysis and Discussion

In this project we devised, conducted and recorded pedestrian movement experiments with the goal of extracting relevant information from the collected trajectories and to help elicitate human behaviour in emergency situations. This chapter presents and discusses the results obtained from following the proposed methodology for achieving the desired aims.

The chapter begins by exposing how the experiments were conducted and then the raw location information collected is presented. The varied visualization tools and methods are depicted in images, followed by discussion of the performance of the data collecting system and the subsequent results of the attempts to further clean, filter and improve the movement data. The single-file experiments were subject of a particular type of analysis, whose outcome is then discussed.

Results of the spatio-temporal analysis and visual inspection are presented next, consisting of some tables summarizing the values of the descriptors and indexes that characterize the trajectories. Several figures illustrating how these values are distributed and evolve along the different areas of the pedestrian facility in question are then presented, followed by a discussion of the results. The chapter ends with the exposition of the knowledge obtained by the usage of data mining techniques on the gathered movement information.

4.1 Pedestrian Experiments

The motion information of pedestrians was recorded from pedestrian movement exercises conducted with up to 30 participants, mostly students, whose average age was 21.4 ± 4.5 years and of mixed gender (Figure 4.1). The experiments were performed in classroom B227 in the Faculty of Engineering of the University of Porto. The area in which the experiments took place was confined within a section of $7m$ x $15m$ of the room, where the only metallic elements present were from the tables that made up barriers for delimiting the track for the experiments.

Figure 4.1: Photograph of participants getting ready for the start of one of the experiments

4.1.1 Data Collection

During the experiments, the application developed to interface with the Ubisense API gathered the location events of the tag in the measurement area and recorded the information in text files. A different file was created for each experiment, and the contents of each of the files following same structure as the following excerpt:

```
1 <Tags>
2 <Tag Name="P30" Date="5/10/2012 12:00:00 AM" Hour="15" Minute="10" Second="27"
      Millisecond="604" x="4.78555774688721" y="-7.63859128952026" z="-0.95" a="
      -0.188959793264351" b="0" c="0" d="0.981984824999599" stderr="
      0.890595674514771" gdop="0" accvalid="1" />
3 <Tag Name="P13" Date="5/10/2012 12:00:00 AM" Hour="15" Minute="10" Second="27"
      Millisecond="604" x="0.939729273319244" y="-9.07961082458496" z="
      -0.950000059604645" a="-0.549598499304991" b="0" c="0" d="0.835428925499771
      " stderr="0.142962396144867" gdop="0" accvalid="1" />
4 <Tag Name="P21" Date="5/10/2012 12:00:00 AM" Hour="15" Minute="10" Second="27"
      Millisecond="619" x="2.24641871452332" y="-10.4371576309204" z="
      -0.950000059604645" a="0.961314486019944" b="0" c="0" d="0.275453188342792"
       stderr="0.354592680931091" gdop="0" accvalid="1" />
5 <Tag Name="P5" Date="5/10/2012 12:00:00 AM" Hour="15" Minute="10" Second="27"
      Millisecond="623" x="5.70134353637695" y="-6.91658163070679" z="-0.95" a="
      0.933159563247838" b="0" c="0" d="0.359462417394509" stderr="
      0.17901137471199" gdop="0" accvalid="1" />
6 </Tags>
```

The UWB system used for data collection reads tags' location asynchronously: a stream of location events is generated over time; each event only contains information about the position of a single tag. Consequently the location of the crowd is updated one pedestrian at a time. Moreover, the system does not ensure a constant frequency of readings for each tag. The mean frequency

of location updates for a single tag in the collected data was $4.74 \pm 1.74Hz$. This compares unfavourably with video collection techniques, where frequencies of $25Hz$ are common and each frame contains data about all pedestrians.

4.1.2 Visualization tool

This simple visualization tool developed for the project takes the previously mentioned log files as input and provides a graphical representation of the tags positions and movement over time in a three-dimensional environment. Figure 4.2 shows the tool in use with data from the one of the experiments. The application can playback the experiments at different time-scales and allows choosing which tags are shown or hidden.

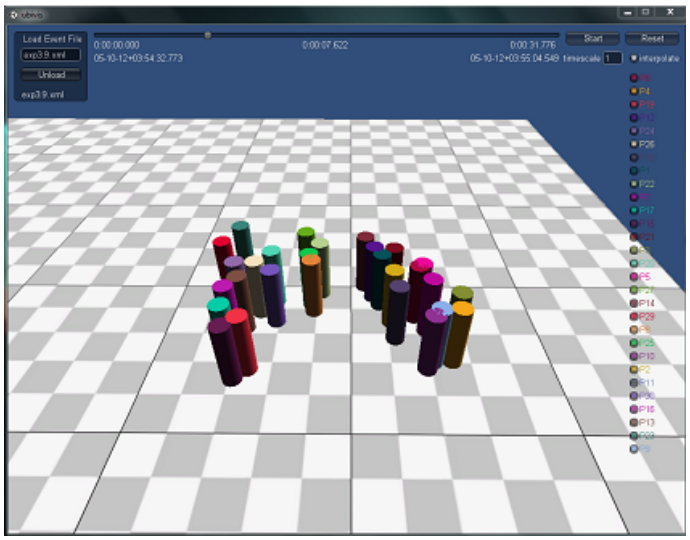


Figure 4.2: Visualization tool

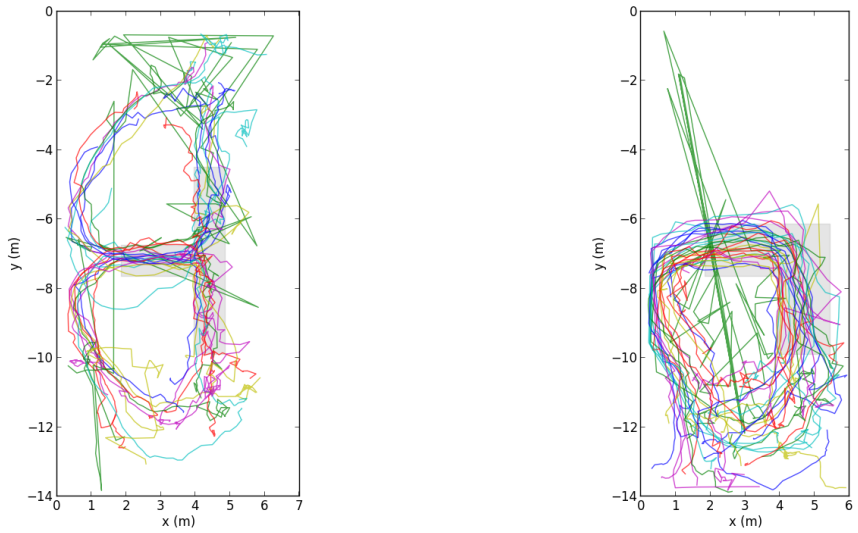
Although this tool presents limited use for analysis, it provided an early insight into the collected data and allowed to understand some limitations of the tracking system. It also allowed the realization that one of the tags used during the experiments - *'P11'* - behaved erratically, and therefore the data associated with that specific tag should be discarded. Another usage was to define the precise instants of the start and end of each experiment.

4.2 Data Filtering and Cleansing

The recorded stream of location events contained spatial and temporal information for the UWB tags used in the experiments. As each tag was carried by a different participant, the trajectory of movement of each pedestrian can be reconstructed from the location events associated with the tag. Trajectories are then stored as a sequence of multidimensional points with two spatial dimensions and another for temporal data.

Assuming piercewise linear movement (i.e. moving along a straight line with constant speed between observations), the graphical representation of the trajectories is simple to obtain and two

examples of such representation for the T-junction and corner experiments are illustrated in Figure 4.3.



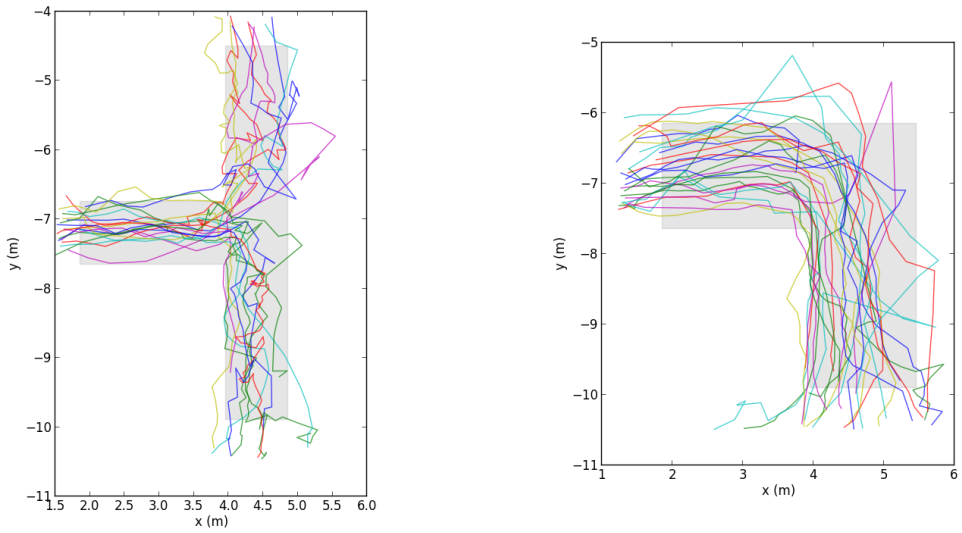
(a) Trajectories for one of the T-Junction experiments (b) Trajectories for one of the Corner experiments

Figure 4.3: Graphical representation of pedestrian trajectories

From the graphical trajectory representation it is immediately obvious that a lot of superfluous information is present in the recorded files and needs to be removed, such as location events from unused tags outside the measurement area that were still tracked during experiments and the data recorded in the instants just before the start and after the end of each experiment duration. It also shows that one of the tags used during the experiments behaved erratically, and therefore should not considered for the purposes of trajectory representation and analysis. Results after discarding bad tags and trimming trajectories are presented in Figure 4.4 for the same experiments as before.

The resulting figures transmit the idea that the collected trajectories appear to be jerky and imprecise. Considerable noise seems to affect some measurements as some trajectories are drawn outside the physically delimited track bounds, where pedestrians would be unable to reach. In particular, some trajectories from the corner experiment (Figure 4.4b) continue moving forward after the corner, moving beyond the track boundaries. This behaviour can in part be explained in the context of the information filter used by the location system framework: position measurements with high uncertainty take a secondary role in comparison with prediction, which uses a rectilinear motion model to predict the position of the tag. Uncertain measurements can therefore cause erroneous trajectories when the tag bearer changes the direction of movement. The trajectories are later promptly corrected when a good measurement occurs.

Although the location system used can, under ideal conditions, achieve an accuracy of up to 20 cm, experiments and test scenarios could only reach sub-meter accuracy [51]. The inaccuracies of the positioning system explain the irregularities in the trajectories, and can be attributed to



(a) Trajectories for one of the T-Junction experiments (b) Trajectories for one of the Corner experiments

Figure 4.4: Graphical representation of pedestrian trimmed trajectories after discarding bad tags

imprecisions in the calibration process, limitations of the information filter used by the location system framework, background noise, the agglomeration of large number of tags in confined areas and the signal attenuation caused by the presence of a large number of test persons.

Such noisy trajectories are not yet adequate for the desired movement analysis, as some of the descriptors of motion such as orientation and acceleration are heavily affected by noise in the position measurements. In order to improve the quality of the recorded trajectories, the data was first filtered to attempt to remove corrupted readings and then smoothed to reduce the noise by smoothing-out short-term irregularity in the data and reveal more clearly the underlying trend.

4.2.1 Data filtering

As mentioned in 3.3.4.1 the filtering algorithm applied to the data was based on Maximum Redundant Distance filter (MRD) of the Douglas Argos-filter algorithm, with some alterations allowing the usage of velocity information instead of spatial distance. The filter works by retaining locations based on spatial redundancy between consecutive locations, therefore removing corrupted readings which are seen as outliers and filtered by the algorithm. Figure 4.5 illustrates the results of applying the filter to a particularly noisy trajectory. The chosen value for the max_v parameter was $1.33 {}^{m}/_{s}$, as obtained from the free velocity in 4.3.1.

Whilst the quality of the filtered trajectory still leaves a lot to be desired, the filter correctly identifies and removes obviously faulty readings.

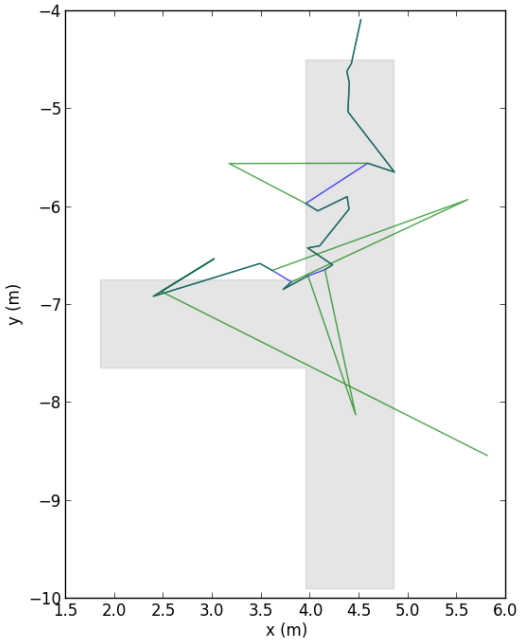


Figure 4.5: Application of the MDR based filter on a particularly noisy trajectory. The original collected trajectory is represented as green and the filtered result as blue.

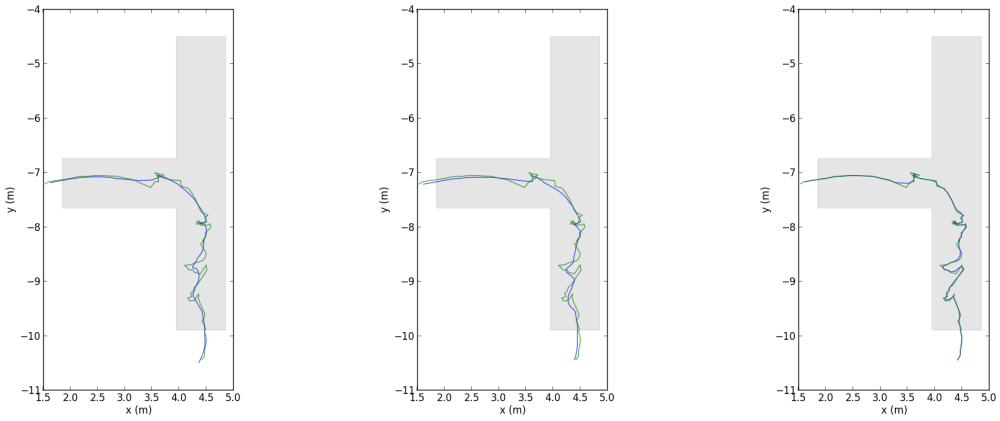
4.2.2 Data smoothing

In this step we attempt to reduce the noise in the data. The irregularities we desire to remove are characterized by short-term disturbances, and therefore we want to reinforce the underlying trend in the data.

As the present work relies on off-line trajectory data, information about the whole trajectory is immediately available, and smoothing techniques can be used to reduce the noise in the readings by improving the location of each measurement using information about previous and subsequent sightings.

Several different approaches to smoothing were trialled; moving average, weighted moving average, exponential and Kalman smoothers were applied to the trajectories. The results of these attempts are depicted in Figures 4.6 and 4.7.

From Figure 4.6 we realise that the Kalman smoother follows the original trajectory too closely, maintaining many of the undesired irregularities, and therefore doesn't perform well in our problem. Although the results of applying the simple moving average, weighted moving average and exponential smoothers seem similar in a two-dimensional representation, the lack of temporal information in both the moving average and exponential smoothers tips the balance in the favour of the usage of the weighted moving average smoother. In the former two smoothers, the position of one observation can be heavily influenced by measurements separated by a significant time interval, whereas usage of weights based on time-differences between observations



(a) Moving average filter (b) Exponential smoother (c) Kalman smoother

Figure 4.6: Different approaches for trajectory smoothing. Green and blue coloured lines represent the trajectory before and after smoothing is applied, respectively.

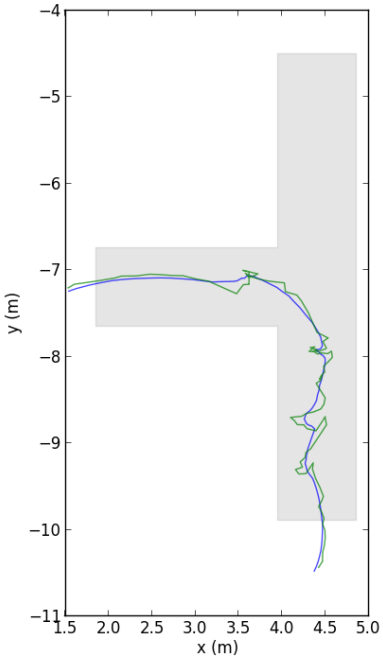


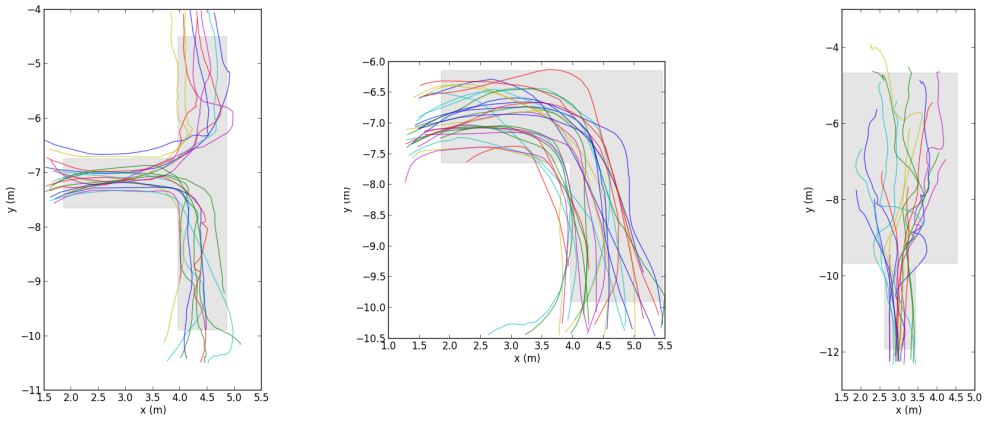
Figure 4.7: Application of the weighted moving average smoother $(k = 9)$ on a trajectory. Green and blue coloured lines represent the trajectory before and after smoothing is applied, respectively.

dampens the effect of distant measurements in the case of the weighted moving average smoother illustrated in Figure 4.7 and detailed in 3.3.4.2. After testing with different window sizes, we found that a value of 9 produced the most appealing results.

4.2.3 Re-sampling

The collected data is sampled at irregular time intervals, and fixed interval re-sampling is required for some of the analysis techniques that were used on the trajectories at a later stage of the project. For an approximate reconstruction of the trajectory, pedestrians are assumed to have a piecewise linear movement - moving along a straight line with constant speed between observations.

After the tasks of cleansing useless information, trimming trajectories, filtering bad readings, smoothing the noise and re-sampling the data are over, the resulting trajectories are exemplified by Figure 4.8.



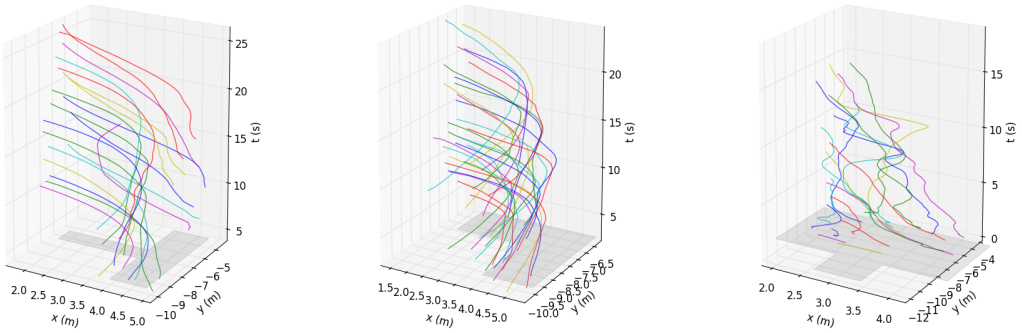
(a) Trajectories for one of the T-Junction experiments

(b) Trajectories for one of the Corner experiments

(c) Trajectories for one of the bottleneck experiments

Figure 4.8: Graphical representation of trajectories after filtering and cleansing

A three-dimensional visualization of the evolution of trajectories over time is also presented in Figure 4.9, where time is associated with the vertical axis and therefore snapshots at different instants can be understood as horizontal planes in the figure.



(a) Trajectories for one of the T-Junction experiments

(b) Trajectories for one of the Corner experiments

(c) Trajectories for one of the bottleneck experiments

Figure 4.9: 3D representation of trajectories

4.3 Initial Analysis

4.3.1 Analysis of the single file experiment

As stated before, the main goal of the single file experiment was to investigate the dependency between the density and velocity of pedestrian movement. Figure 4.10 shows the extracted trajectories for this scenario.

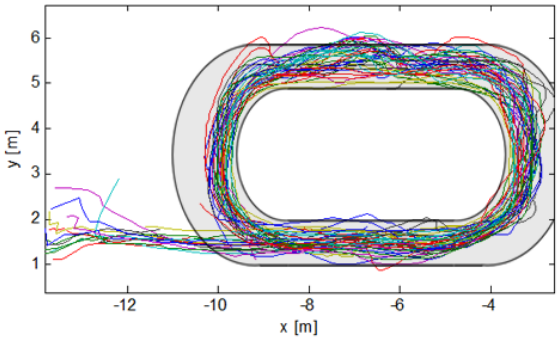


Figure 4.10: Extracted trajectories for the single file scenario

Towards this goal, only a straight section with length $l_m = 2m$ (as seen in Figure 4.11) was considered to determine the density-velocity relation of pedestrian movement. Entrance (t_{en}) and exit (t_{ex}) times were recorded for each pedestrian crossing the entrance (x_{en}) and exit (x_{ex}) of this section. From these times, both the average velocity of each crossing i as well as the number of persons inside the measurement section $N(t)$ at each instant t were obtained.

From the runs with a single pedestrian on the track, the free velocity $v_{free} = 1.33 \pm 0.13 m/s$, was obtained, which matches quite well with the value from literature ($1.34 m/s$) [66].

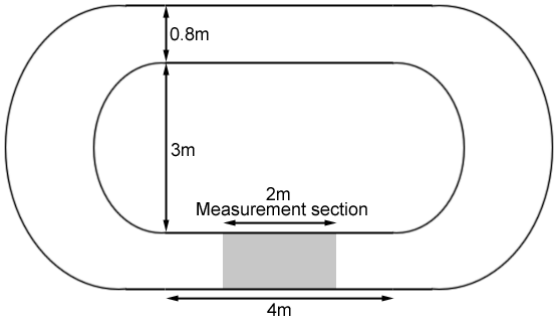


Figure 4.11: Experimental setup for the single file scenario

Figure 4.12 shows the evolution of the crossings' speed and density in the measurement section over the whole duration of the run with 20 pedestrians. As the pedestrian density in the measurement area at a given time t is obtained by simply dividing the section length by the number of pedestrians inside it at t, it will jump between discrete values over the duration of the experiment, and is represented by the black line the figure. Tick blue lines, whose length indicates the time interval a pedestrian is inside the measurement section, represent the mean velocity of the crossing.

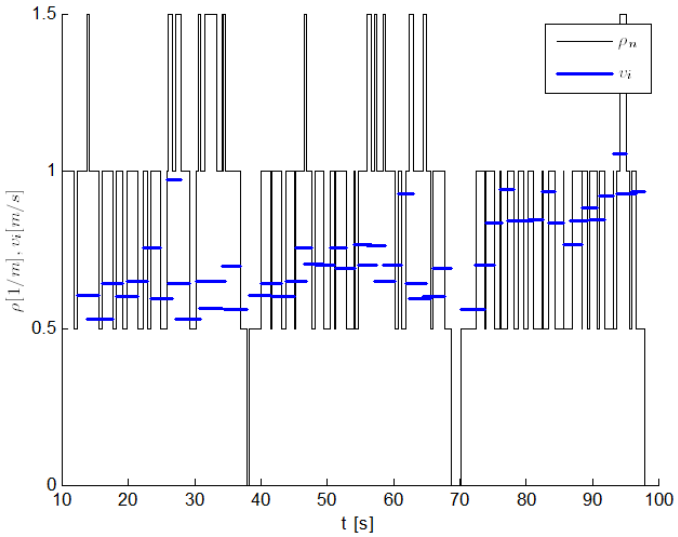


Figure 4.12: Evolution of crossing speed v_i and density in the measurement section ρ_n over the duration of the experiment composed of 20 participants. Tick blue lines, whose length indicates the time interval a pedestrian is inside the measurement section, represent the mean velocity of the crossing.

The density assigned to each pedestrian crossing the measurement section is determined as the mean value of density in the section during the crossing. A graphical representation of the velocity (v_i) - density (ρ_i) pairs of each crossing is presented in Figure 4.13. This representation is known as the *fundamental diagram of pedestrian movement*.

In comparison with the diagrams obtained from similar experiments [20, 21] (Figure 4.14), where data was collected manually from video recordings, similar values for density and velocity are found for runs with the same number of participants. However, our diagram is more disperse as

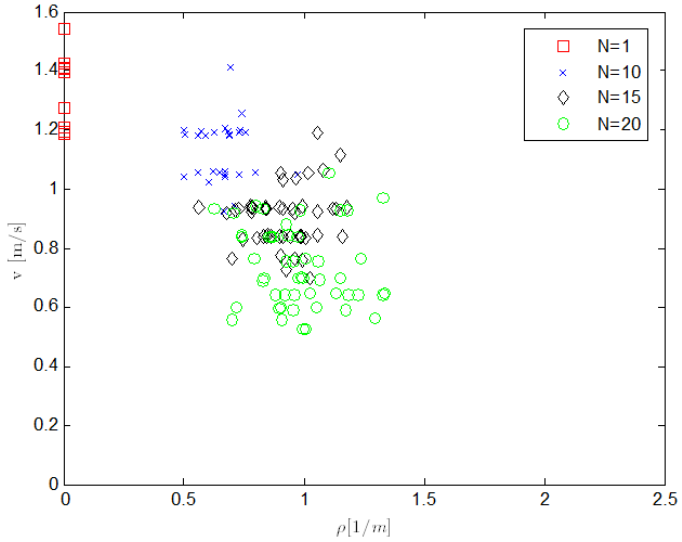


Figure 4.13: Relation between density and velocity (fundamental diagram) in the single file scenario for the runs with 1, 10, 15 and 20 participants.

a result of the limited precision of the UWB system. No data is presented for values of density over $1.2 {}^{1}/_{m}$ because when the experiment was performed no more than 20 voluntaries were present.

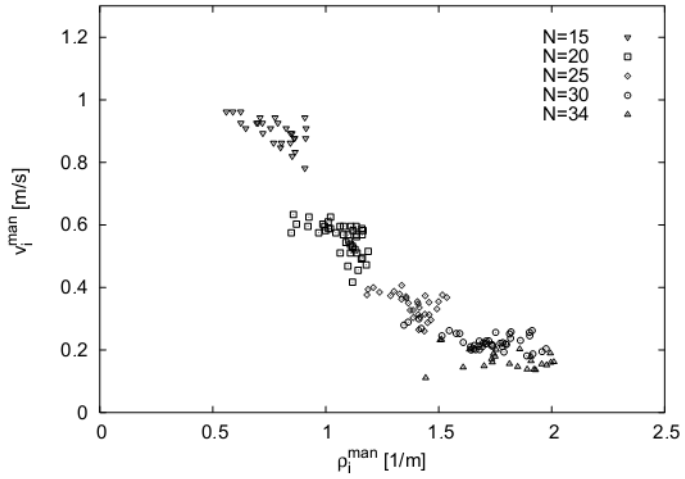


Figure 4.14: Relation between density and velocity from similar experiments. From [20].

4.4 Spatio-temporal Analysis and Visual Inspection

4.4.1 Aggregate trajectory characterization

Aggregate trajectory characterization aimed at providing values quantifying trajectories as a whole. In each of the experiments, for each of the participating pedestrian's trajectories, the following values that hold information about that whole trajectory were obtained: displacement, evacuation

time, trajectory length, average velocity, straightness index, directional mean and circular variance.

Table 4.1 shows the results of quantifying trajectories as a whole for the data collected from the first run of the T-Junction experiments as an example of the results obtained in this step.

Table 4.1: Aggregate motion descriptors of trajectories for one of the T-Junction experiments

		Mean	SD	Min	Max
Trajectory	Evacuation Time (sec)	9.24	2.44	5	13.8
	Average Velocity (m/s)	0.602	0.131	0.388	0.855
	Path Length (m)	5.38	0.596	4.02	7.11
	Displacement (m)	3.44	0.849	0.322	4.28
	Straightness Index	0.638	0.153	0.0723	0.795
	Average Orientation (rad)	0.503	2.13	-2.34	2.55
	Circular Dispersion	0.985	0.00557	0.975	0.995

The relations that arise between some of motion descriptors can be analysed with the help of Table 4.2, which lists the correlation matrix of trajectories motion descriptors. Observation of the table shows that some obvious correlations are clearly visible; for example positive correlation among motion descriptors describing sinuosity of path including straightness index, straight length and circular dispersion; and the negative correlation between evacuation time and average velocity of the movement.

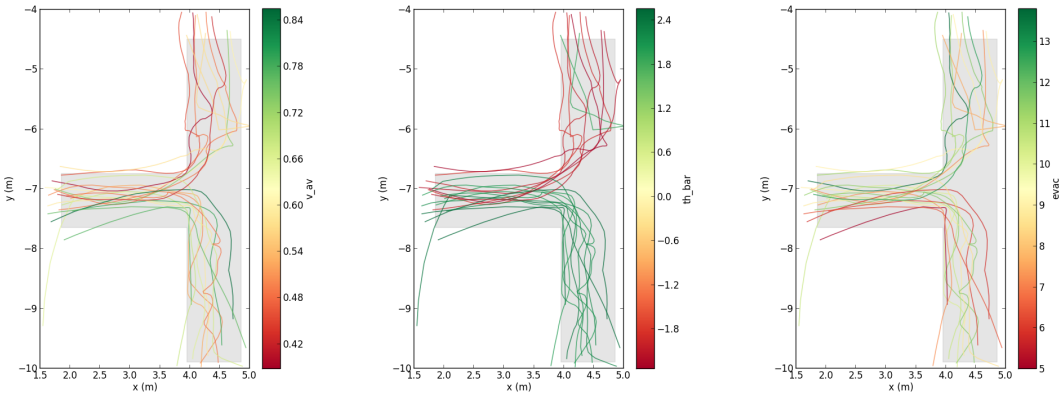
Table 4.2: Correlation matrix of aggregate motion descriptors of trajectories for one of the T-Junction experiments

	Evac. Time	Av. Velocity	Path Length	Displacement	St. Index	Av. Orientation	Circ. Disp.
Evac. Time	1.0000						
Av. Velocity	-0.9291	1.0000					
Path Length	0.5621	-0.2797	1.0000				
Displacement	0.2914	-0.1229	0.3844	1.0000			
St. Index	0.0756	0.0262	0.0685	0.9388	1.0000		
Av. Orientation	-0.4963	0.5291	-0.1028	-0.3178	-0.2712	1.0000	
Circ. Disp.	0.7473	-0.7826	0.4404	-0.3009	-0.5383	-0.3150	1.0000

For all the runs in the bottleneck experiment the pedestrians occupy the same initial position and only the exit width is changed. As expected, evacuation times decrease $(12.6, 11.4, 8.37s)$ and average velocities increase $(0.585, 0.708, 0.844^{m}/_{s})$ as the exit width is increased $(0.9, 1.0, 1.4m)$.

The same can't be said about the corner experiments, however. The runs with: equal entry and exit widths of $1.5m$; $1.5m$ entry and $0.9m$ exit; and equal entry and exit widths of $0.9m$, feature evacuation times of $6.81, 9.55$ and $9.6s$ and average velocities of $0.88, 0.712$ and $0.799^{m}/_{s}$ respectively. These results imply that, although evacuation time only shows dependency with the exit width, velocity varies with entry width as well.

Two dimensional maps of trajectories, where each trajectory from one of the T-junction experiments is coloured by the corresponding value of a set of aggregate motion descriptors are presented in Figure 4.15. From a initial visual analysis, the behavioural information transmitted by these representations is low, although it can be noted that in the case of the average orientation figure a clear partition between the trajectories originated from the movement of the pedestrians that come from different entrances to the Junction is immediately clear (Figure 4.15b).



(a) Average trajectory velocity (b) Average orientation (c) Evacuation time

Figure 4.15: Example of T-Junction trajectories coloured by aggregate descriptors

Besides the obvious observation that pedestrians whose initial position is close to the exit have lower evacuation times, Figure 4.16 also shows that the evacuation time of pedestrians originating from a central area is lower that whose of the periphery.

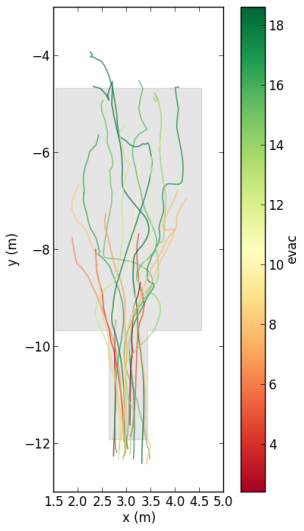


Figure 4.16: Trajectories coloured by evacuation time for one of the bottleneck experiments.

Aggregate characterization is not very handy at providing insight into the details of how properties of pedestrian flow such as densities and velocities are spread in facilities, or how they vary from one facility to another. However, they seem to be very important in realizing how evacuation times change depending on the pedestrians initial positions, path and exit choices during emergencies.

Some of the computed descriptors such as the straightness index and circular dispersion can provide insight into the ability of evacuation participants to efficiently find egress routes. Even simple properties such as evacuation times are important benchmarks of evacuation exercises and important for building planners.

4.4.2 Segmented trajectory characterization

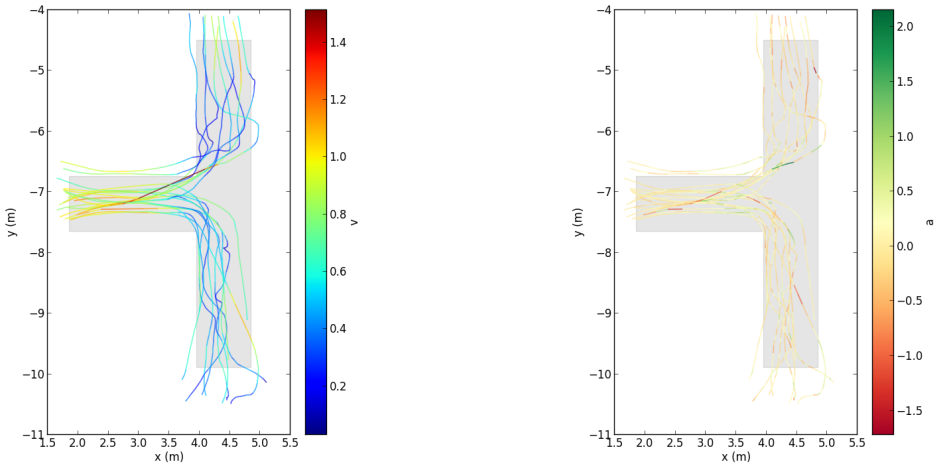
Segmented trajectory characterization aimed at further improving the already existent description of movement. Since trajectories are assumed to be represented as a set of points with an associated spatial location and time instant, each trajectory can also be imagined as a the sequence of segments connecting two consecutive points.

A vector of new features was then computed for each segment of the trajectories and stored alongside the location and temporal data. The following descriptors that quantify information about each segment of the trajectory were obtained: displacement, velocity, acceleration, orientation and density. Information about the displacement, velocity and acceleration along the x and y axes were also calculated.

Representations of trajectories in which each of the segments that constitute a trajectory is coloured according to the value of descriptors associated with motion data from that segment are depicted in the figures that follow. Figure 4.17 illustrates the two-dimensional maps of trajectories for of the T-junction experiments obtained when considering the average velocity and acceleration of segments, respectively.

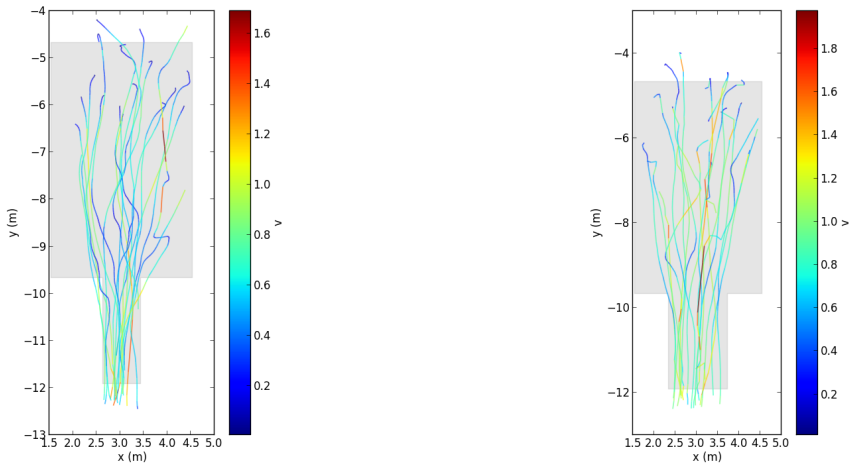
The information conveyed by the segment coloured maps allows distinction of different values of descriptors such as velocity for separate parts of each trajectory. This permits a glimpse on how the characteristics of movement are distributed throughout the different zones of the experiment area. For instance, Figure 4.17a transmits the idea that the pedestrians movement speed is slow as they approach the junction, but faster after the merging of streams. The acceleration map reinforces the same idea, as the centre of the junction features a large number of segments with positive acceleration, whilst the corridors leading up to it show the opposite.

Average segment velocity coloured trajectories for the bottleneck experiment with widths of 0.9 and 1.4m are depicted in Figure 4.18. Is is apparent that the lower velocity, higher congestion areas appear before the bottleneck, with pedestrians moving faster once they find themselves already inside the narrow section of the facility. The figure also showcases the effect of the width of the bottleneck on the variation of velocity: the colour of segments in figure 4.18a features a wider range that that of figure 4.18b, meaning a wider variation of velocity - caused by the congestion of a narrower exit.



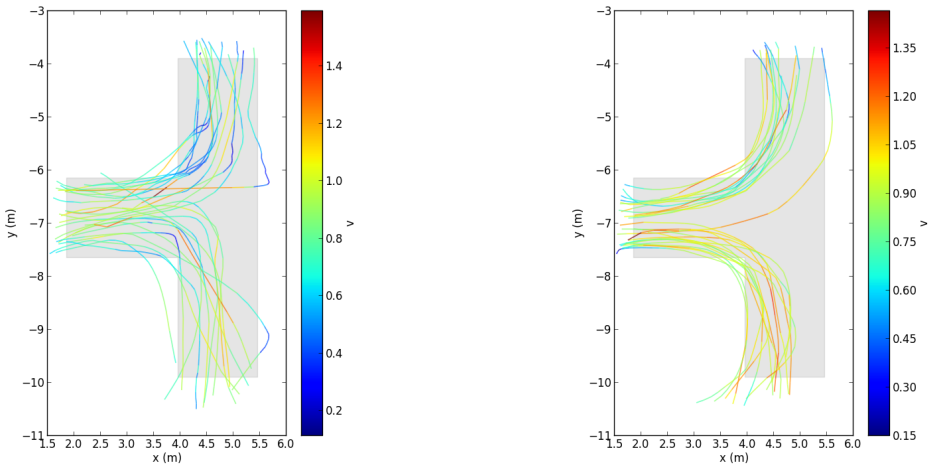
(a) Average segment velocity (b) Average segment aceleration

Figure 4.17: Trajectories coloured according to the value of motion descriptors associated with each segment



(a) Bottleneck width 0.9m (b) Bottleneck width 1.4m

Figure 4.18: Trajectories coloured according to the value of average segment velocity for different runs of the bottleneck experiment



(a) Merging of flows

(b) Splitting of flow

Figure 4.19: Trajectories coloured according to the value of average segment velocity for different flow directions in the T-junction experiments

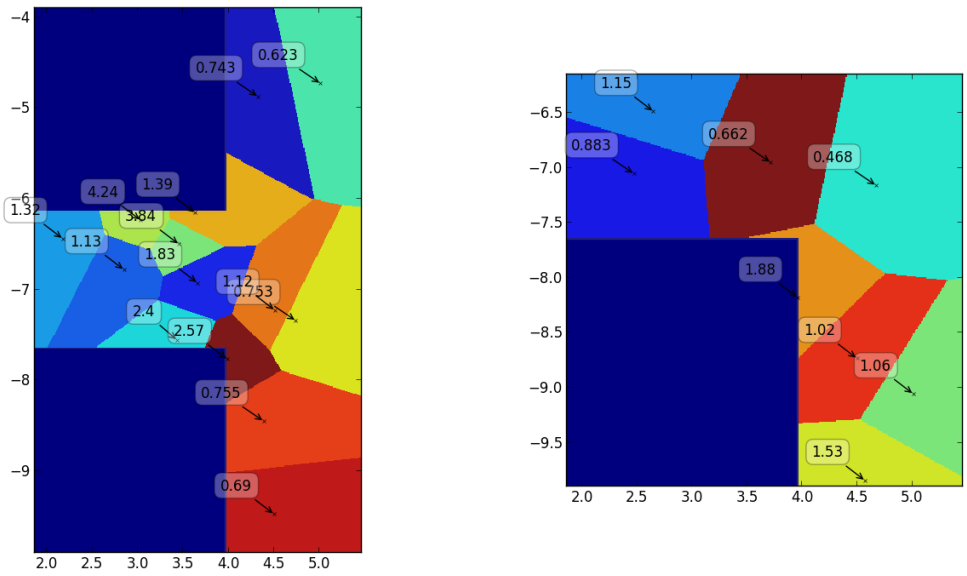
Figure 4.19 illustrating average segment velocity for different flow directions in a T-junction like facility transmits the idea that, for the same overall number of pedestrians, a splitting flow is characterized by higher movement velocity than that of merging flows. The motion also seems more regular in the splitting flow figure (4.19b), with reduced velocity variance and pedestrians moving orderly and closer to the walls and resulting in shorter paths.

4.4.2.1 Density obtained through the Voronoi partition of space

In order to assign densities at different temporal instants to the different zones of the experiment area, a Voronoi partition of space, where the position of the tags is input as seeds to the method, is performed. For each instant t, the cells obtained by the method represent the *personal area* occupied by each pedestrian at that instant. This allows measuring the density in any point in space as the reciprocal of the area of the cell that the point belongs to. Figure 4.20 represents the partition of space in different experiment areas according to the Voronoi method at a certain instant. The crosses represent the position of pedestrians and the labels show the density to assigned to each.

The density in the T-junction (Figure 4.20a) is not homogeneous and a higher density region appears near the junction. The lowest density region is located at a triangle-shaped area where the left and right branches begin to merge. The densities in the branches (before the merging region) are not uniform and are higher over the inner side, especially near the corners. Similar conclusions are extracted by looking at Figure 4.20b, referencing the distribution of density at a certain time instant t, in the corner experiment.

Density in the bottleneck experiment takes the highest value just before the narrowing, decreasing as the distance to it increases as illustrated in Figure 4.21. A reduction the values of



(a) T-junction (b) Corner

Figure 4.20: Partition of space in the experiment areas according to the Voronoi method. The crosses show the positions of pedestrians at the time and the labels indicate the assigned density. Density is calculated as the reciprocal of the area occupied by each cell.

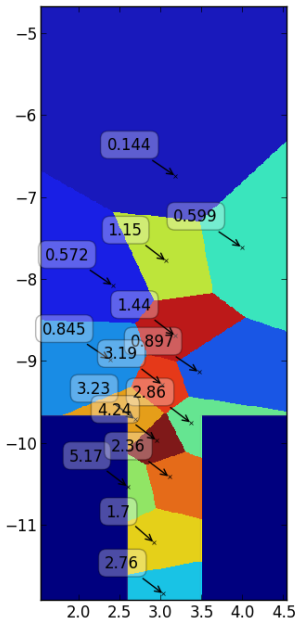


Figure 4.21: Partition of space in the bottleneck experiment according to the Voronoi method.

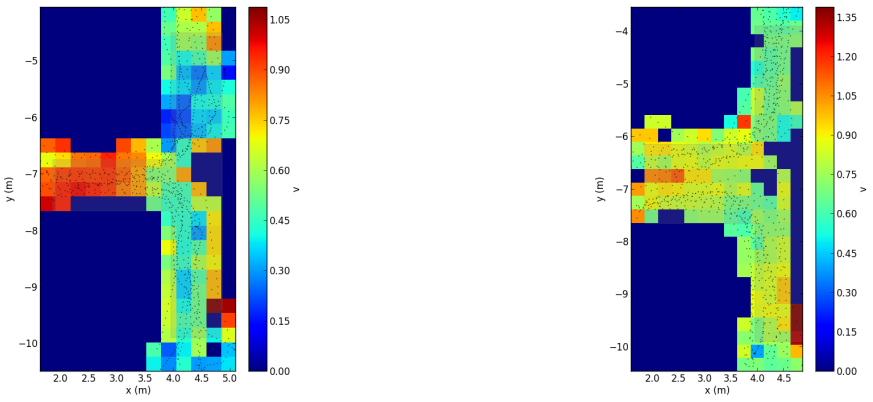
density is also strongly related with the increasing distance along the direction perpendicular to the narrow passage.

4.4.3 Visualization through histograms and KDE maps

With the sole exception of the Voronoi based representation of pedestrian density data, all the results presented until now only contained information about trajectories and the space directly traversed by pedestrians. In an attempt to provide results and information for every point of space in the facilities, we extrapolate information from neighbour locations that already contain the data.

In the first attempt, we divide the space along a two-dimensional grid, and assume that the value of a certain motion descriptor of each cell of the grid is the average of the values of that same descriptor from known locations in that cell. This idea is analogous to a 2D-histogram, and is illustrated in Figure 4.22. The figures clearly show that, for different runs of the T-junction experiment with different exit widths, the velocity is more regular and homogeneous along the whole experiment area in the case where the flows merge into the larger corridor of Figure 4.22b.

The conclusions found by analysing Figure 4.22a are the same discussed preciously on a similar analysis of Figure 4.17a: movement is slow as pedestrians approach the junction, but faster after the merging of streams. However, relation seems much more clear in the newly obtained histogram figure.



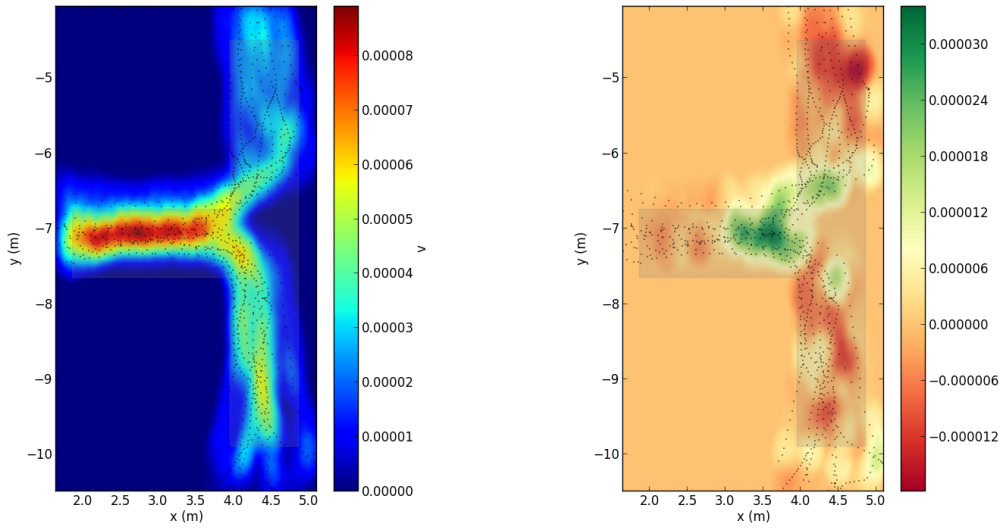
(a) Entry and exit widths are the same and equal to 0.9m

(b) Entry width of 0.9m on each side and common exit with 1.5m of width

Figure 4.22: Histogram of velocity for T-junction experiments with merging flows and different entry and exit widths

Still, the histogram approach has some limitations. Finding appropriate grid size is problematic and sometimes no good solution can be found. Wide grid sizes generalize too much and valuable information is lost. Narrow grid sizes risk making many cells empty of known descriptor locations and therefore without a way to extrapolate information. This problem is exacerbated in the facilities where some areas are very populated and others not visited very often.

On the second attempt, we try a non-parametric way to estimate the probability density function of the descriptor. Kernel density estimation allows us to make inferences about the whole experiment are based on a finite data sample. Two-dimensional maps of kernel density estimates for velocity and acceleration in the T-junction experiments can be visualized in Figure 4.23.



(a) Average segment velocity

(b) Average segment acceleration

Figure 4.23: KDE map of trajectories descriptors

The kernel density maps provide qualitative information about the distribution of the descriptors. Although lacking the important quantitative information, these maps make up to it by presenting information in a smoother manner, and also providing information about zones not covered previously (e.g. zones were not crossed by trajectories) due to its probabilistic nature. Again, the intersection where pedestrians meet is shown to be the bottleneck of the evacuation as the velocity is lower before it and higher after (Figure 4.23a), which correspond to areas of deceleration and acceleration respectively (Figure 4.23b).

4.5 Spatio-temporal Data Mining

4.5.1 Trajectory partition

Trajectory partitioning aims to differentiate contrasting behaviour within a trajectory. Trajectory partitions are sub-trajectories that are characterized by different behavioural characteristics. To achieve this goal, the partitioning algorithm looks for the characteristic points of a trajectory - points where the behaviour of the trajectory changes rapidly. Each resulting trajectory partition is represented by a line segment between two consecutive characteristic points.

The partitioning algorithm implemented in this project is called TRACLUS and was introduced in [64]. Figure 4.24 illustrates a pedestrian trajectory and the respective segments between characteristic points produced by the partitioning algorithm.

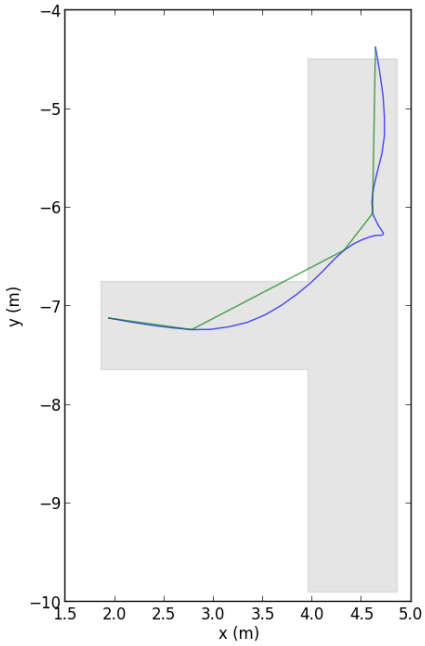


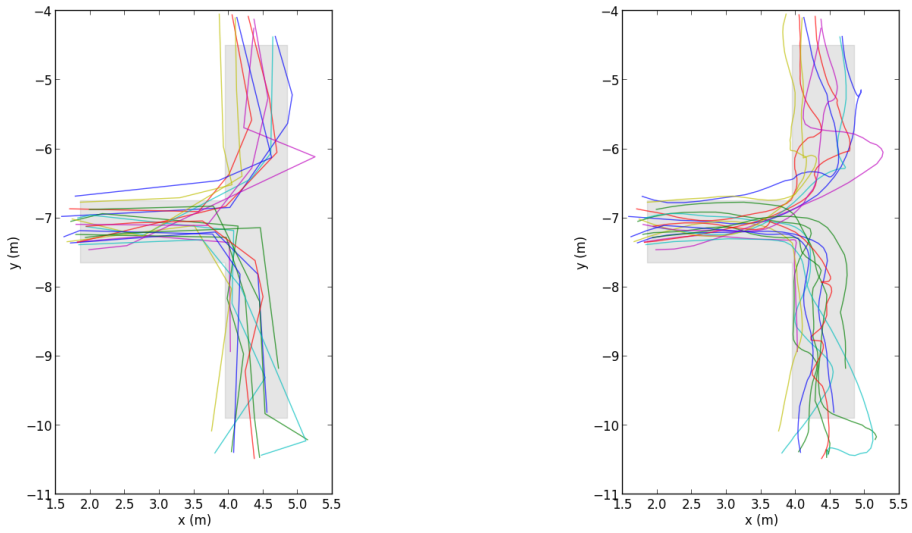
Figure 4.24: Results of partitioning a trajectory. Represented in blue colour is the original trajectory and the segments between characteristic points resulting from trajectory partitioning are plotted in green.

Figure 4.25a depicts the segments resulting from partitioning all the trajectories for one of the T-junction experiments. Inspection of the figure reveals that, although much simpler that the original trajectories, the partition segments still retain most of the fundamental geometrical information of the original trajectories, whilst eliminating minute details that, due to the imprecisions in the data collection system, have little significance.

The trajectory partitioning approach adopted in this project essentially considers only the directionality of the movement. It is, thus, useful to find behavioural changes that are accompanied by changes in the motion direction. This approach is not optimal, as important behavioural changes occur in situations such as congestion due to bottlenecks, where the geometry of trajectories is mostly unaffected but properties of movement such velocities are deeply affected.

In his thesis [67], Nara proposes a different partitioning approach that relies on the assumption that in many situations pedestrians change of behaviour involves moments of stopping or staying. This approach is based on a distance threshold: if the distance of each segment is less that a threshold the segment is assigned as stay.

However, due to the short term nature of the data collected in this project, and the imprecisions and noise that affect the data - even after extensive smoothing and treatment, properties such as



(a) Segments resulting from trajectory partitioning

(b) Original trajectories

Figure 4.25: Results of partitioning multiple trajectories

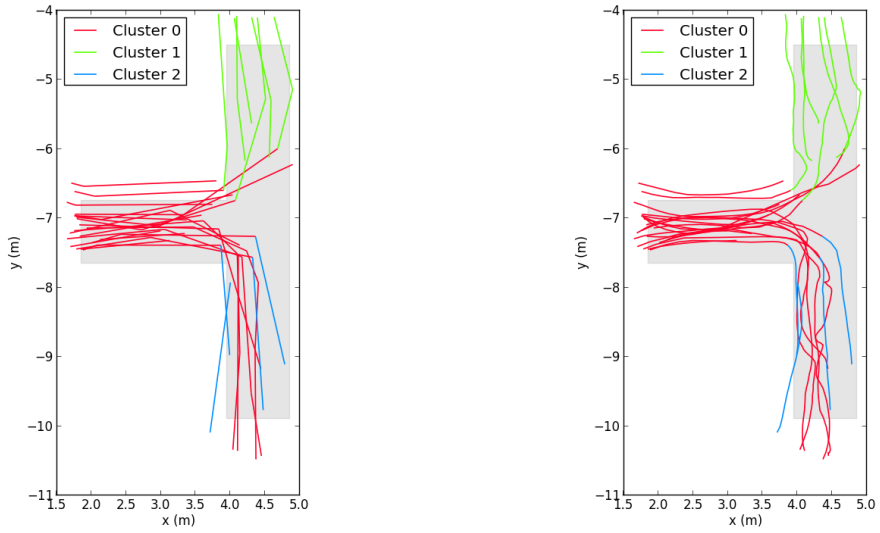
velocities still maintain large fluctuations that impede the use of this alternative approach.

4.5.2 Segment clustering

After the different portions of trajectories with different behavioural characteristics are separated, it is time to discover which sub-trajectories share common movement patterns and group them into clusters for further study and characterization of the behaviour singled out by each cluster.

Figure 4.26 shows the different segment clusters, and the corresponding sub-trajectories, found in the data of one of the runs of the T-junction experiment. Three clusters are identified: each featuring segments and sub-trajectories from different branches of the facility (with the exception of some segments from the lower branch appearing the in the mid branch cluster). The results are satisfactory and mostly in line with what we expected to obtain, because, as with the trajectory partitioning approach, the clustering algorithm adopted in this project is mostly based on the geometry of the segments.

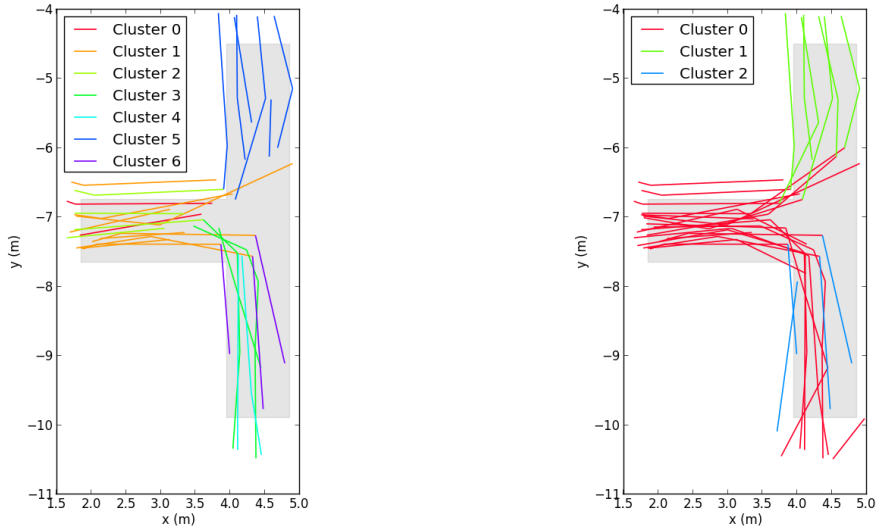
Lee et al. [64] also produce a heuristic for parameter value selection, that in the case of the T-junction scenario returns the values $\varepsilon = 1.2$ and $MinLns = 3$ for parameters. Figure 4.27a illustrates the clusters resulting by using the parameter values provided by the heuristic. Usage of these values results on increased number of clusters, some of which appear very similar. In comparison, the clusters identified in Figure 4.27b, where the values for parameters were chosen by manual tuning, clearly separate sub-trajectories in the different zones of the facility: clusters 0 and 2 groups sub-trajectories in left and right branches and cluster 1 represents movement after merging. The number of segments not assigned to any cluster is also inferior in the case where the parameters were manually tuned.



(a) Segments

(b) Sub-trajectories

Figure 4.26: Results of clustering



(a) Parameters $\varepsilon = 1.2$ and $MinLns = 3$ provided by the heuristic

(b) Parameters $\varepsilon = 1.55$ and $MinLns = 3$ found through manual tuning

Figure 4.27: Effect of different parameters on clustering

Results also show that the discovered clusters are highly dependant on the chosen parameters, and that the optimal parameters change with the geometry of the facility, therefore requiring constant turning for different facilities. These limitations are caused by the geometry based approach to the trajectory segmentation and clustering.

4.5.3 Cluster characterization

This final step involves gathering aggregate information about the sub-trajectories that make up each cluster. The characterization is analogous to that applied to the whole trajectories in 4.4.1. The information created in this step aims to allow statistical modelling of the clusters as a means for human movement model elicitation.

One of the sets of sub-trajectories produced by partitioning trajectories from the the first run of the T-Junction experiments based on its different behaviours, and by clustering similarly behaved sub-trajectories is graphically illustrated in Figure 4.28.

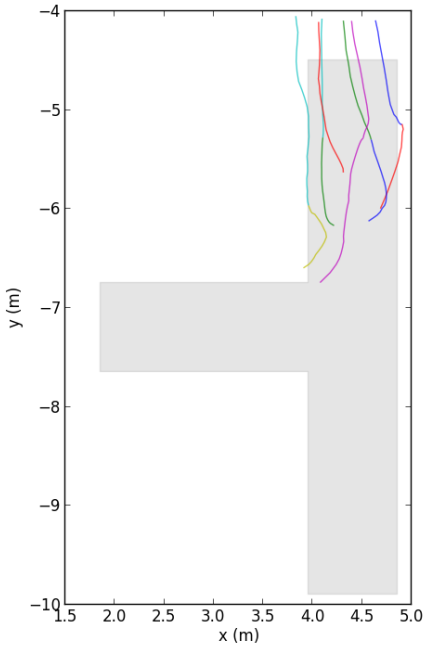


Figure 4.28: Sub-trajectories for the cluster 1 of one of the T-Junction experiments.

Table 4.3 shows the results of quantifying the previously mentioned sub-trajectories as an example of the results obtained by cluster characterization.

Statistical information such as the mean and standard deviation of the average orientation (pointing downwards in this example), velocity and displacement could be sufficient for creating artificial trajectories similar to those that belong to the cluster in analysis. The same procedure

Table 4.3: Aggregate motion descriptors of sub-trajectories for the cluster 1 of one of the T-Junction experiments

		Mean	SD	Min	Max
Trajectory	Evacuation Time (sec)	3.62	1.02	2.4	6
	Average Velocity (m/s)	0.342	0.0815	0.223	0.456
	Path Length (m)	1.26	0.33	0.847	1.97
	Displacement (m)	1.18	0.357	0.635	1.92
	Straightness Index	0.919	0.069	0.75	0.975
	Average Orientation (rad)	-1.52	0.202	-1.89	-1.16
	Circular Dispersion	0.947	0.0131	0.929	0.969

applied to all the clusters found in a complete set of trajectories could help recreate the same trajectories, thus acting as a rudimentary modelling tool for the experiment from which the trajectories were extracted.

Chapter 5

Conclusions

5.1 Final remarks

Given the objectives of the project: study and improve techniques of tracking human movement using UWB RFID tags; conduct experiments with volunteers and record tracked trajectories; extract relevant information from the recorded trajectories; use data mining and inference techniques to help elicitate human behaviour, and the findings provided by the analysis of the conducted experiments, we conclude that this research met the initial goals.

Several experiments of pedestrian movement in restricted areas were devised and conducted, from which trajectories were recorded by a UWB RDIF tracking system. Regarding the first two objectives, we concluded that the Ubisense RTLS lacks the fine spatial precision that would be desired to make extensive quantitative analysis of motion at the scales relevant for characteristics of movement such as velocities and accelerations. We can also reason that although the system has a complex and time-consuming setup and installation, its usage and handle is relatively accessible.

Usage of this technology presents several advantages over traditional data collection techniques, expanding the breath of possible scenarios for experiments, such as situations of limited visibility for which data is inexistent. Compared with the usual method for tracking of people trajectories, video recordings, UWB allows use in narrower spaces, facilities with lower ceilings and areas with line of sight restrictions. Other aspect is related with data collection. UWB based systems record the coordinates of position for each tag directly whereas video recordings must be later analysed and processed in order to extract positions / trajectories. Each tag carried by pedestrians is uniquely identifiable, allowing individual tracking and thus enabling the investigation of the behaviour and effect of specific individuals like child, elderly or people with mobility impairments on the crowd dynamics.

On the other hand, the UWB system used has a lower sample rate: about 5 Hz in the case of UWB versus 25 Hz with video. Also, video techniques present synchronized results whilst UWB does not. Moreover, technical limitations like the lack of fine spatial precision make it not optimal for extracting microscopic properties of traffic, such as velocities and densities at a disaggregated level.

In conclusion, and when comparing the drawbacks with the advantages, UWB techniques for human trajectory extraction seems a viable approach. It is particularly suitable for scenarios where video is less applicable and pedestrian fine positioning is not an issue, such as for macroscopic analysis (e.g. egress time, path choice and behaviour scrutiny).

Concerning the information extracted from the trajectories, we found that the system limitations like the lack of fine spatial precision became a problem when trying to extract detailed information, for example, for performing quantitative analysis. Attempts to improve the quality of the data by filtering and smoothing proved to be challenging tasks in this project.

When it comes to the third objective — extract relevant information from the recorded trajectories, it is also important to mention that adequate choice of parameters is fundamental to properly characterize and describe the motion behaviour patent in the trajectory data.

The computed descriptors of movement, combined with the exploration of several different visualization techniques, provided insight into some of the processes of crowd congestion and emergency egress. The results from this step capture and describe the spatio-temporal behaviour of crowds, an important feature of evacuation dynamics. Such results can be used for improving facility design as well evacuation route planning.

On the subject of using data mining techniques for discovering hidden trends and knowledge from the trajectory dataset, we discovered that out approach was valid, as we were able to extract patterns and its characterising variables such as velocities, orientations and tortuosity indexes. Although the patterns were also easily identifiable by with visual analysis of the trajectories' characteristics, the data-mining method provides an automatic alternative, that can be used on massive datasets as no human input is required. It also proved that, by using machine learning techniques, it can adapt to different situations and scenarios that alternative conventional methods might not.

5.2 Future Developments

Although the present study makes a contribution to the database of pedestrian movement knowledge, it also intends to incite further research on the matter of the usage of UWB based tracking systems. This is a promising technology that widens the range of possible test scenarios for which there is scarce data such as low visibility situations, so common during evacuations due to fire or during the night. As this is a relatively new topic, there are still some subjects that remain unexplored, thus this promotes an opportunity for further investigation.

Some limitations of the current study are related to the lack of the desired accuracy and precision from the tracking system used, which brings consequences related to the ability to discern smaller nuances in the behavioural patterns. Therefore, future studies that build on these results and suppress its limitations are recommended and include better filtering and smoothing approaches and sensor fusion.

The spatio-temporal analysis and visualization techniques that made the basis of the methodological approach of this project can also be applied to other sources of pedestrian motion information, such as simulations in virtual environments and data collected with the help of other technologies.

Important problems that were not optimally solved in this project still have plenty of room for improvement. Examples of such problems are the limitations of the directional-based trajectory partition algorithm and the density-based clustering of segments of trajectories. Description of subsections of trajectories by feature vectors, dimensionality reduction by principal component analysis and the the application of more conventional clustering algorithms are some interesting approaches that could be applied to the data-mining problematic of the project.

Future work will also focus on the analysis of other experiments using different scenarios, following the same methodology. The results of this work will be used to validate and calibrate behaviour models in the "mSPEED" framework. This will constitute a unique tool for agent-based "Modelling and Simulation of Pedestrian Emergent Evacuation Dynamics" under development at LIACC.

References

[1] João E Almeida, Rosaldo Rosseti, and António Leça Coelho. Crowd Simulation Modeling Applied to Emergency and Evacuation Simulations using Multi-Agent Systems. In *DSIE'11 - 6th Doctoral Symposium on Informatics Engineering*, 6th Doctoral Symposium on Informatics Engineering, DSIE'11, Porto, pages 93–104, 2011. URL: http://paginas.fe.up.pt/~prodei/dsie11/images/pdfs/s3-1.pdf.

[2] J. E. Almeida, Z. Kokkinogenis, and R. J. R. Rosseti. NetLogo Implementation of an Evacuation Scenario. In *WISA 2012 (Fourth Workshop on Intelligent Systems and Applications)*, Madrid, 2012.

[3] Andreas Schadschneider, Wolfram Klingsch, Hubert Kluepfel, Tobias Kretz, Christian Rogsch, and Armin Seyfried. Evacuation Dynamics: Empirical Results, Modeling and Applications. *Encyclopedia of Complexity and System Science*, page 57, 2008. URL: http://arxiv.org/abs/0802.1620.

[4] Dirk Helbing, Péter Molnár, Illés J Farkas, and Kai Bolay. Self-organizing pedestrian movement. *Environment and Planning B Planning and Design*, 28(3):361–383, 2001. URL: http://www.envplan.com/abstract.cgi?id=b2697, doi:10.1068/b2697.

[5] M Boltes, A Seyfried, B Steffen, and A Schadschneider. Automatic Extraction of Pedestrian Trajectories from Video Recordings. *Pedestrian and Evacuation Dynamics 2008*, pages 1–13, 2008. URL: http://www.springerlink.com/index/n420218717p18252.pdf, doi:10.1007/978-3-642-04504-2.

[6] B D Hankin and R A Wright. *Passenger Flow in Subways*, volume 9. Operational Research Quarterly, 1958. URL: http://www.jstor.org/stable/3006732.

[7] S J Older. Movement of pedestrians on footways in shopping streets. *Traffic Engineering and Control*, 10(4):160–163, 1968. URL: http://www.citeulike.org/user/MehdiMoussaid/article/4372359.

[8] Ulrich Weidmann. Transporttechnik der Fußgänger - Transporttechnische Eigenschaften des Fußgängerverkehrs (Literaturauswertung). In *Schriftenreihe des IVT 90*. Institut fuer Verkehrsplanung und Transportsysteme, 1992.

[9] Núria Pelechano and Ali Malkawi. Comparison of crowd simulation for building evacuation and an alternative approach. *Building Simulation 2007 Vols 13 Proceedings*, pages 1514–1521, 2007.

[10] S P Hoogendoorn, P H L Bovy, and W Daamen. Microscopic Pedestrian Wayfinding and Dynamics Modelling. In M Schreckenberg and S D Sharma, editors, *Pedestrian and Evacuation Dynamics*, pages 124–154. Springer, 2002. URL: http://en.scientificcommons.org/1416895.

[11] Winnie Daamen. *Modelling Passenger Flows in Public Transport Facilities*. PhD thesis, TU Delft, 2004. URL: http://www.narcis.nl/publication/RecordID/oai:tudelft.nl:uuid:e65fb66c-1e55-4e63-8c49-5199d40f60e1.

[12] Michael Scott Ramming. *Network Knowledge and Route Choice*. PhD thesis, Massachusetts Institute of Technology, 2002. URL: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Network+Knowledge+and+Route+Choice#0.

[13] D Helbing and A Johansson. Pedestrian, Crowd and Evacuation Dynamics 1. *Media*, 16(4):6476–6495, 2010. URL: http://www.springerlink.com/index/10.1007/978-3-642-04504-2, doi:10.1007/978-3-642-04504-2.

[14] P D Navin and R J Wheeler. *Pedestrian flow characteristics*, volume 10. Traffic Engineering and Control, 1968. URL: http://tris.trb.org/view.aspx?id=116327.

[15] J J Fruin. *Pedestrian Planning and Design*, volume 77. Elevator World, Inc., Educational Services Division, 1971. URL: http://www.ncbi.nlm.nih.gov/pubmed/22061941, doi:10.1016/j.meatsci.2007.05.004.

[16] A Seyfried, M Boltes, J Kähler, W Klingsch, A Portz, T Rupprecht, A Schadschneider, B Steffen, and A Winkens. Enhanced empirical data for the fundamental diagram and the flow through bottlenecks. *Review Literature And Arts Of The Americas*, page 12, 2008. URL: http://arxiv.org/abs/0810.1945.

[17] Jun Zhang, Wolfram Klingsch, Andreas Schadschneider, and Armin Seyfried. Transitions in pedestrian fundamental diagrams of straight corridors and T-junctions. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(06):17, 2011. URL: http://arxiv.org/abs/1102.4766.

[18] Dirk Helbing. Derivation of a Fundamental Diagram for Urban Traffic Flow. *European Physical Journal B*, 70(2):229–241, 2008. URL: http://arxiv.org/abs/0807.1843.

[19] Florian Siebel and Wolfram Mauser. On the fundamental diagram of traffic flow. *Transportation Research*, 66(June):7, 2005. URL: http://arxiv.org/abs/cond-mat/0503290.

[20] Armin Seyfried, Bernhard Steffen, Wolfram Klingsch, and Maik Boltes. The Fundamental Diagram of Pedestrian Movement Revisited. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(10):13, 2005. URL: http://arxiv.org/abs/physics/0506170.

[21] Ujjal Chattaraj, Armin Seyfried, and Partha Chakroborty. Comparison of Pedestrian Fundamental Diagram Across Cultures. *Advances in Complex Systems*, 12(3):12, 2009. URL: http://arxiv.org/abs/0903.0149.

[22] Dirk Helbing and Peter Molnar. Social Force Model for Pedestrian Dynamics. *Physical Review E*, 51(5):4282–4286, 1998. URL: http://arxiv.org/abs/cond-mat/9805244.

[23] Bernhard Steffen and Armin Seyfried. Methods for measuring pedestrian density, flow, speed and direction with minimal scatter. *Physica A: Statistical Mechanics and its Applications*, 389(9):16, 2009. URL: http://arxiv.org/abs/0911.2165.

[24] J Zhang, W Klingsch, T Rupprecht, and A Schadschneider. Empirical study of turning and merging of pedestrians streams in T-junction. 2010. arXiv:1112.5299v1.

[25] Serge P Hoogendoorn and W Daamen. Pedestrian Behavior at Bottlenecks. *Transportation Science*, 39(2):147–159, 2005. URL: http://transci.journal.informs.org/cgi/doi/10.1287/trsc.1040.0102, doi:10.1287/trsc.1040.0102.

[26] D Helbing, L Buzna, A Johansson, and T Werner. Self-organized pedestrian crowd dynamics: Experiments, simulations, and design solutions. *Transportation Science*, 39(1):1–24, 2005. URL: http://discovery.ucl.ac.uk/150154/.

[27] Tobias Kretz, Anna Grünebohm, and Michael Schreckenberg. Experimental study of pedestrian flow through a bottleneck. *Journal of Statistical Mechanics: Theory and Experiment*, 2006(10):P10014–P10014, 2006. URL: http://arxiv.org/abs/physics/0610077.

[28] EM Cepolina and N Tyler. Understanding Capacity Drop for designing pedestrian environments. *Sixth International Walk21 Conference*, 2005. URL: http://discovery.ucl.ac.uk/1412/.

[29] Armin Seyfried, Tobias Rupprecht, Oliver Passon, Bernhard Steffen, Wolfram Klingsch, and Maik Boltes. New insights into pedestrian flow through bottlenecks. *Transportation Science*, 43(3):16, 2007. URL: http://arxiv.org/abs/physics/0702004.

[30] S P Hoogendoorn, W Daamen, and P H L Bovy. Microscopic Pedestrian Traffic Data Collection and Analysis by Walking Experiments: Behaviour at Bottlenecks. In E R Galea, editor, *Pedestrian and Evacuation Dynamics 03*, pages 89–100. CMS Press, London, 2003.

[31] Armin Seyfried, Bernhard Steffen, Andreas Winkens, Tobias Rupprecht, M Boltes, and W Klingsch. Empirical data for pedestrian flow through bottlenecks. *Traffic and Granular Flow'07*, (1):189–199, 2009. URL: http://www.springerlink.com/index/w6063p3g5792117t.pdf.

[32] Dirk Helbing, Illes Farkas, and Tamas Vicsek. Simulating Dynamical Features of Escape Panic. *Nature*, 407(6803):487–490, 2000. URL: http://arxiv.org/abs/cond-mat/0009448.

[33] Dirk Helbing, Illes Farkas, and Tamas Vicsek. Freezing by Heating in a Driven Mesoscopic System. *Physical Review Letters*, 84(6):1240–1243, 1999. URL: http://arxiv.org/abs/cond-mat/9904326.

[34] Dirk Helbing and Pratik Mukerji. Crowd Disasters as Systemic Failures: Analysis of the Love Parade Disaster. 2012. arXiv:arXiv:1206.5856v1.

[35] H W Hamacher and S A Tjandra. Mathematical Modelling of Evacuation Problems : A State of Art. *Pedestrian and Evacuation Dynamics*, 24(24):227–266, 2001. URL: http://www.forschungsplattform.de/zentral/download/berichte/bericht24.pdf.

[36] E R Galea. Simulating Evacuation and Circulation in Planes, Trains, Buildings and Ships using the EXODUS software. In M Schreckenberg and S D Sharma, editors, *Pedestrian and Evacuation Dynamics*, pages 203–226. Springer, 2002.

[37] S P Hoogendoorn and P H L Bovy. State-of-the-art of vehicular traffic flow modelling. *Proceedings of the I MECH E Part I Journal of Systems Control in Engineer*, 215(4):283–303, 2001. URL: http://www.catchword.com/rpsv/cgi-bin/cgi?ini=xref&body=linker&reqdoi=10.1243/0959651011541120, doi:10.1243/0959651011541120.

[38] K Teknomo. Microscopic pedestrian flow characteristics: Development of an image processing data collection and simulation model. *PhD thesis Tohoku University Japan Sendai*, (March), 2002. URL: http://people.revoledu.com/kardi/publication/Dissertation.pdf.

[39] T Klupfel H Meyer-Konig. Characteristics of the PedGo Software for Crowd Movement and Egress Simulation. In *2nd International Conference in Pedestrian and Evacuation Dynamics PED*, pages 331–340, 2003.

[40] L F Henderson. The statistics of crowd fluids., 1971. URL: http://www.ncbi.nlm.nih.gov/pubmed/16059256.

[41] Victor J Blue and Jeffrey L Adler. Cellular Automata Microsimulation of Bidirectional Pedestrian Flows. *Transportation Research Record*, 1678(1):135–141, 1999. URL: http://trb.metapress.com/openurl.asp?genre=article&id=doi:10.3141/1678-17, doi:10.3141/1678-17.

[42] S. Sharma and S. Gifford. Using RFID to Evaluate Evacuation Behavior Models. *NAFIPS 2005 - 2005 Annual Meeting of the North American Fuzzy Information Processing Society*, pages 804–808. URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1548643, doi:10.1109/NAFIPS.2005.1548643.

[43] Da Zhang, Feng Xia, Zhuo Yang, Lin Yao, and Wenhong Zhao. Localization Technologies for Indoor Human Tracking. *2010 5th International Conference on Future Information Technology*, (60903153):1–6, 2010. URL: http://arxiv.org/abs/1003.1833.

[44] Faranak Nekoogar. Introduction to Ultra-Wideband Communications. In *UltraWideband Communications Fundamentals and Applications*, chapter 1, pages 1–44. Prentice Hall, 2005. URL: http://www.informit.com/articles/article.aspx?p=433381.

[45] J A Corrales, F A Candelas, and F Torres. Hybrid tracking of human operators using IMU/UWB data fusion by a Kalman filter. *Proceedings of the 3rd international conference on Human robot interaction HRI 08*, page 193, 2008. URL: http://portal.acm.org/citation.cfm?doid=1349822.1349848, doi:10.1145/1349822.1349848.

[46] Usama M Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. From data mining to knowledge discovery: an overview. In U M Fayyad, G Piatetsky-Shapiro, P Smyth, and R Uthurusamy, editors, *Advances in Knowledge Discovery and Data Mining*, chapter 1, pages 1–34. American Association for Artificial Intelligence, 1996. URL: http://portal.acm.org/citation.cfm?id=257942.

[47] H Michael Chung and Paul Gray. Special section: data mining. *Journal of Management Information Systems*, 16(1):11–16, 1999.

[48] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. *Introduction to Data Mining, (First Edition)*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2005.

[49] H J Miller and J Han. *Geographic Data Mining and Knowledge Discovery*. Chapman & Hall/CRC Data Mining and Knowledge Discovery Series. Taylor & Francis, 2003. URL: http://books.google.pt/books?id=fJTD26iYxpkC.

[50] Xiaobai Yao. Research issues in spatio-temporal data mining. *White paper UCGIS*, 2003.

[51] Jose Xavier, Pedro Henriques Abreu, Luis Paulo Reis, and Marcelo Petry. Location and automatic trajectory calculation of mobile objects using radio frequency identification, 2011.

[52] P Steggles and S Gschwind. The Ubisense smart space platform. In *Adjunct Proceedings of the Third International Conference on Pervasive Computing*, volume 191, pages 73–76. ACM Press, 2005. URL: http://www.pervasive.ifi.lmu.de/adjunct-proceedings/demo/p073-076.pdf, doi:10.1145/313451.313476.

[53] J Cadman. Deploying commercial location-aware systems. In *Proceedings of the 2003 Workshop on LocationAware Computing*, pages 4–6, 2003.

[54] Mitja Luštrek, Boštjan Kaluža, Erik Dovgan, Bogdan Pogorelc, and Matjaž Gams. Behavior analysis based on coordinates of body tags. In Manfred Tscheligi, Boris De Ruyter, Panos Markopoulus, Reiner Wichert, Thomas Mirlacher, Alexander Meschterjakov, and Wolfgang Reitberger, editors, *Ambient Intelligence*, volume 5859 of *Lecture notes in computer science*, pages 14–23. Springer Berlin / Heidelberg, 2009. URL: http://dis.ijs.si/bostjan/papers/BehaviorAnalysisBasedonCoordinatesofBodyTags.pdf, doi:10.1007/978-3-642-05408-2_2.

[55] Maik Boltes, Jun Zhang, Armin Seyfried, and Bernhard Steffen. T-junction: Experiments, trajectory collection, and analysis, 2011. URL: http://scholar.google.com/scholar?hl=en&q=author:Boltes,m.+flow+OR+pedestrian&btnG=Search&as_sdt=0,5&as_ylo=&as_vis=0, doi:10.1109/ICCVW.2011.6130238.

[56] Ubisense. How to configure filters using Location Engine Config. URL: http://eval.ubisense.net/howto/LocationEngineFilters_article/LocationEngineFilters.html.

[57] C Douglas David, Rolf Weinzierl, Sarah Davidson, Roland Kays, Martin Wikelski, and Gil Bohrer. Moderating Argos location errors in animal tracking data. *Methods in Ecology and Evolution*, 2012, 2012. doi:10.1111/j.2041-210X.2012.00245.x.

[58] Paulo J A L Almeida, Marcus V Vieira, Maja Kajin, German Forero-Medina, and Rui Cerqueira. Indices of movement behaviour: conceptual background, effects of scale and location errors. *Zoologia Curitiba Impresso*, 27(5):674–680, 2010. URL: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1984-46702010000500002&lng=en&nrm=iso&tlng=en, doi:10.1590/S1984-46702010000500002.

[59] E Batschelet. *Circular Statistics in Biology*, volume 24 of *Mathematics in Biology*. Academic Press, 1981. URL: http://www.jstor.org/stable/1267831?origin=crossref, doi:10.2307/1267831.

[60] Abigail L McCarthy, Selina Heppell, Francois Royer, Carla Freitas, and Thomas Dellinger. Identification of likely foraging habitat of pelagic loggerhead sea turtles (Caretta caretta) in the North Atlantic through analysis of telemetry track sinuosity. *Progress in Oceanography*, 86(1-2):224–231, 2010. URL: http://linkinghub.elsevier.com/retrieve/pii/S0079661110000431, doi:10.1016/j.pocean.2010.04.009.

[61] G L Gaile and J E Burt. *Directional statistics*. Concepts and Techniques in Modern Geography Series. Geo Abstracts, University of East Anglia, 1980. URL: http://books.google.pt/books?id=KA12QgAACAAJ.

[62] Emanuel Parzen. On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics*, 33(3):pp. 1065–1076, 1962. URL: http://www.jstor.org/stable/2237880.

[63] B W Silverman. *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability. Chapman and Hall, 1986. URL: http://books.google.pt/books?id=e-xsrjsL7WkC.

[64] Jae-Gil Lee, Jiawei Han, and Kyu-Young Whang. Trajectory clustering: a partition-and-group framework. *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, pages 593–604, 2007. URL: http://portal.acm.org/citation.cfm?id=1247546, doi:10.1145/1247480.1247546.

[65] Martin Ester, Hans-peter Kriegel, Jörg Sander, and Xiaowei Xu. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *Computer*, 1996(6):226–231, 1996. URL: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:A+Density-Based+Algorithm+for+Discovering+Clusters+in+Large+Spatial+Databases+with+Noise#0, doi:10.1.1.71.1980.

[66] Erica D Kuligowski, Richard Peacock, and Bryan L Hoskins. A Review of Building Evacuation Models , 2nd Edition, 2010. URL: http://www.nist.gov/manuscript-publication-search.cfm?pub_id=906951.

[67] Atsushi Nara. *Developing a Cohesive Space-time Information Framework for Analyzing Movement Trajectories in Real and Simulated Environments*. PhD thesis, Arizona State University, 2011.