A DISCRETE CHOICE MODELING FRAMEWORK FOR PEDESTRIAN WALKING BEHAVIOR WITH APPLICATION TO HUMAN TRACKING IN VIDEO SEQUENCES

THÈSE N^{O} 3382 (2005)

PRÉSENTÉE À LA FACULTÉ SCIENCES ET TECHNIQUES DE L'INGÉNIEUR

Institut de traitement des signaux

SECTION DE GÉNIE ÉLECTRIQUE ET ÉLECTRONIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Gianluca ANTONINI

Laurea di Dottore in Ingegneria delle Telecomunicazioni, Università degli Studi di Siena, Italie
et de nationalité italienne

acceptée sur proposition du jury:

Prof. J.-P. Thiran, Dr M. Bierlaire, directeurs de thèse
Prof. M. Ben Akiva, rapporteur
Dr A. Cavallaro, rapporteur
Prof. R. Siegwart, rapporteur

Lausanne, EPFL
2006

Preface

I was very excited when i have started to work on my PhD project, because of the cross-disciplinary nature of the subject. The original idea was actually to investigate and analyze pedestrian dynamics from a behavioral point of view, and use this kind of *prior* information for multi-object tracking in video sequences. Human behavior modeling represents a very complex and challenging task, undertaken by classical disciplines such as social sciences and psychology. Specific *instances* of human behavior have been studied for a long time in economics (consumer behavior), transportation science (driver behavior), cognitive sciences (individuals' perception) and artificial intelligence (agent-based systems). The first important step in this work was the choice of a behavioral paradigm. I decided for the use of *rational behavior*, implemented by means of discrete choice models, more specifically random utility models.

Pedestrian modeling and simulation represent by *per se* recent research fields. Growing interest is in the context of crowd evacuation management and panic situation analysis, for obvious reasons. Capturing the behavior of individuals in *normal* situations is instead important for architecture, urban planning, land use, marketing, traffic operations. The second step of this work was to make a bridge between discrete choice analysis and pedestrian modeling and simulation.

Recent applications and research directions in the image processing and computer vision communities (video surveillance, scene analysis, activity recognition) have triggered the interest in model-based approaches. Different sources of *prior* information are modeled and integrated into algorithms for fundamental problems, such as image segmentation, object detection and tracking. This context represents the main motivation for the third step in this work, where I have tried to combine behavioral models for pedestrian dynamics as a source of *prior* information for multi-object tracking.

Every tracking system suffers from a bias in the estimation of the number of targets. This fact represents the motivation for the last part of this thesis. A post-processing of the tracking output is performed, using clustering techniques, in order to reduce the bias in the targets' number estimation.

The work has been appealing but not always simple. The methodological choices have been made in the belief they were representing the best solution for that specific problem at that time. I have given the preference to a cross-disciplinary approach, highlighting the integration between the different parts, rather than the single part.

Acknowledgements

I am grateful to my two supervisors, Prof. Jean Philippe Thiran and Dr. Michel Bierlaire. They have been precious from both a technical and human point of view. I want to thank particularly Michel Bierlaire to have "opened" the MIT door and both of them to provide me with th necessary financial aid to support that project. I am grateful to Prof. Moshe Ben-Akiva to accepted me at the Intelligent Transportation System Group at MIT. It has been a wonderful experience, given me a lot of motivations and ideas to finish the modeling part. A special thanks is for Carmine Gioia, met at MIT, and who is become one of my best friends. Our discussions have been precious for my work as well as for my life in general. I want to thank all my friends in Lausanne and my colleagues, especially Matteo Sorci, always present in the last 13 years. A very special thank to my parents and my sister, supporting me as usual in any situation. Finally, i want to dedicate this work to my new family, Marjorie and Annamè, for their unconditioned and always present love.

Abstract

Intelligent Transportation Systems (ITS) have triggered important research activities in the context of behavioral dynamics. Several new models and simulators for driving and travel behaviors, along with new integrated systems to manage various elements of ITS, have been proposed in the past decades. In this context, less attention has been given to pedestrian modeling and simulation. In 2001, the first international conference on Pedestrian and Evacuation Dynamics took place in Duisburg, Germany, showing the recent, growing interest in pedestrian simulation and modeling in the scientific community. The ability of predicting the movements of pedestrians is valuable indeed in many contexts. Architects are interested in understanding how individuals move into buildings to find out optimality criteria for space design. Transport engineers face the problem of integration of transportation facilities, with particular emphasis on safety issues for pedestrians. Recent tragic events have increased the interest for automatic video surveillance systems, able to monitoring pedestrian flows in public spaces, throwing alarms when abnormal behaviors occur. In this spirit, it is important to define mathematical models based on specific (and context-dependent) behavioral assumptions, tested by means of proper statistical methods. Data collection for pedestrian dynamics is particularly difficult and few models presented in literature have been calibrated and validated on real datasets.

Pedestrian behavior can be modelled at various scales. This work addresses the problem of pedestrian *walking* behavior modeling, interpreting the walking process as a sequence of choices over time. People are assumed to be *rational* decision makers. They are involved in the process of choosing their next position in the surrounding space, as a function of their kinematic characteristics and reacting to the presence of other individuals. We choose a mathematical framework based on *discrete choice analysis*, which provides a set of well founded econometric tools to model disaggregate phenomena.

The pedestrian model is applied in a computer vision application, namely detection and tracking of pedestrians in video sequences. A methodology to integrate behavioral and image-based information is proposed. The result of this approach is a *dynamic detection* of the individuals in the video sequence. We do not make a clear cut between detection and tracking, which are rather thought as inter-operating procedures, in order to generate a set of hypothetical pedestrian trajectories, evaluated with the proposed model, exploiting both dynamic and behavioral information. The main advantage applying such methodology is given by the fact that the standard target detection/recognition step is bypassed, reducing the complexity of the system, with a consistent gain in computational time. On the other hand, the price to pay as a consequence for the simple initialization procedure is the overestimation of the number of targets. In order to reduce the bias in the targets' number estimation, a comparative study between different approaches, based on clustering techniques, is proposed.

Versione abbreviata

Negli ultimi anni la ricerca nei Sistemi di Trasporto Intelligenti (ITS) ha visto lo sviluppo di numerosi lavori nel campo delle dinamiche comportamentali. Nuovi modelli e simulatori per lo studio del traffico, insieme allo sviluppo di sistemi integrati per la gestione dei vari elementi dell'ITS, sono stati proposti negli ultimi decenni. In questo contesto, pochi sforzi sono stati fatti nella modellizzazione e simulazione del traffico pedonale. La prima conferenza internazionale sullo studio delle dinamiche pedonali e di evacuazione (PED 2001) ha avuto luogo a Duisburg, Germania, nel 2001, a testimonianza del crescente interesse della comunita' scientifica per l'argomento. L'abilita' nel predirre i movimenti pedonali e' valutabile in vari contesti. In architettura, i progettisti sono interessati a capire come gli individui si muovono all'interno di edifici e spazi in generale, per trarre criteri di ottimalita' nella progettazione degli spazi stessi. In ingegneria dei trasporti si ha il problema dell'integrazione delle infrastrutture per traffico pedonale e veicolare, dove particolare enfasi e' dedicata a problemi di sicurezza per i pedoni. I tragici avvenimenti degli ultimi anni hanno aumentato l'interesse per sistemi automatizzati di video-sorveglianza, per il monitoraggio dei pedoni in spazi publici, e la possibilita' di avere sistemi di allarme in caso di comportamenti anormali. In questo contesto, e' importante definire modelli matematici basati su specifiche ipotesi comportamentali, testati con l'ausilio di metodi statistici. La raccolta di dati per il traffico pedonale e' particolarmente difficile e pochi dei modelli presentati in letteratura sono stati calibrati e validati su dati reali.

Il comportamento pedonale puo essere modellato su scale differenti. In questa tesi il problema affrontato e' quello della modellizzazione di agenti pedonali dinamici, *walking behavior*. Le traiettorie degli agenti nello spazio sono interpretate come una sequenza di scelte nel tempo. Gli individui sono assunti come unita' decisionali *razionali*, coinvolti nella scelta di *dove muovere il prossimo passo* nello spazio circostante, in funzione dei propri parametri cinematici e in reazione alla presenza e movimento di altri pedoni. Il framework matematico usato e' quello dei modelli di scelta discreti (DCM), il quale fornisce un insieme di tecniche econometriche per la modellizzazione di comportamenti disaggregati. Il modello sviluppato e' calibrato su dati reali, raccolti con *tracking* manuale di pedoni su sequenze video.

Il modello e' stato poi applicato ad un problema di *computer vision*, piu precisamente alla detezione e tracking di pedoni in sequenze video. Una metodologia per l'integrazione dell'informazione comportamentale e quella basata sull'immagine e' proposta. Il risultato e' una *detezione dinamica* degli individui nella sequenza video. Tracking e detezione sono considerati come due processi interagenti, piuttosto che come due parti separate. Un insieme di traiettorie pedonali ipotetiche e' generato con tecniche standard di object-tracking. Il modello e' in seguito utilizzato per filtrare l'insieme delle traiettorie che piu si avvicinano al comportamento pedonale, cosi come descritto dal modello stesso. Il vantaggio principale di questo approccio consiste nella detezione dinamica. Complessi algoritmi

di segmentazione e riconoscimento di oggetti sull'immagine sono cosi evitati, riducendo la complessita' del sistema, con conseguente guadagno in termini computazionali. La semplicita' usata per la parte di inizializzazione ha come conseguenza un fenomeno di detezione multipla dei pedoni nella sequenza video. Il sistema di tracking stima per eccesso il vero numero di individui presenti nella scena. Per ridurre questo errore, uno studio comparativo basato su tecniche di clustering applicate alle traiettorie generate dal sistema di detezione e tracking e' stato proposto.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 A twofold motivation

In the last years several domains of engineering have seen the increase of interest in pedestrian behavior modeling. The huge amount of people in city centres, shopping malls, public infrastructures, such as rail stations and airports, triggers the interest of different research communities in understanding and describing behavior of individuals in such situations. Getting insights in pedestrian modeling is important for public space designers, dealing with space constraints and congestion problems, for transportation researchers, dealing with facilities integration, demand modeling and safety issues. Recently (and unfortunately) tragic events have also triggered the interest in evacuation dynamics and modeling. Individual dynamics, in all these situations, are dictated by behavioral issues. Defining models based on precise behavioral assumptions, well founded mathematical frameworks, and supported by a real data calibration step, represents at the same time a first challenge and motivation for this work. In this spirit, inspiration is taken from transportation science, where sophisticated modeling techniques, aiming to reproduce driver and travel behaviors, have been borrowed from microeconomics and econometrics, and widely used in different contexts. I try here to extend part of these ideas to the pedestrian case.

A multitude of recording devices are placed almost everywhere, providing hours of collected videos. Scientists in computer vision are interested in extracting information from image sequences, aiming to perform automatic tasks, such as pedestrian detection and tracking, activity monitoring and recognition and scene analysis in general. Detection and tracking of individuals in video sequences is one of the most important tools for many applications, from video surveillance to scene analysis in general. The final goal in such systems consists in the recognition and evaluation of human activities.* The adopted solution is, in most of the cases, a hierarchical one. Detection and tracking represent *low level* tasks, based on the image information, while *higher level* tasks, such as activity recognition and evaluation, are treated in post-processing steps. This point of view is summarized in Figure 1.1. Given the complexity of real scenarios, and the difficulties making a bridge between low level and high level tasks, in the last years the research community moved towards integrated methods. Several sources of information are combined together, trying to overcome the limits imposed by a pure image-based methodology. This context represents the second motivation for

*We refer implicitly to scenes where only pedestrian dynamics and the related activities are of interest

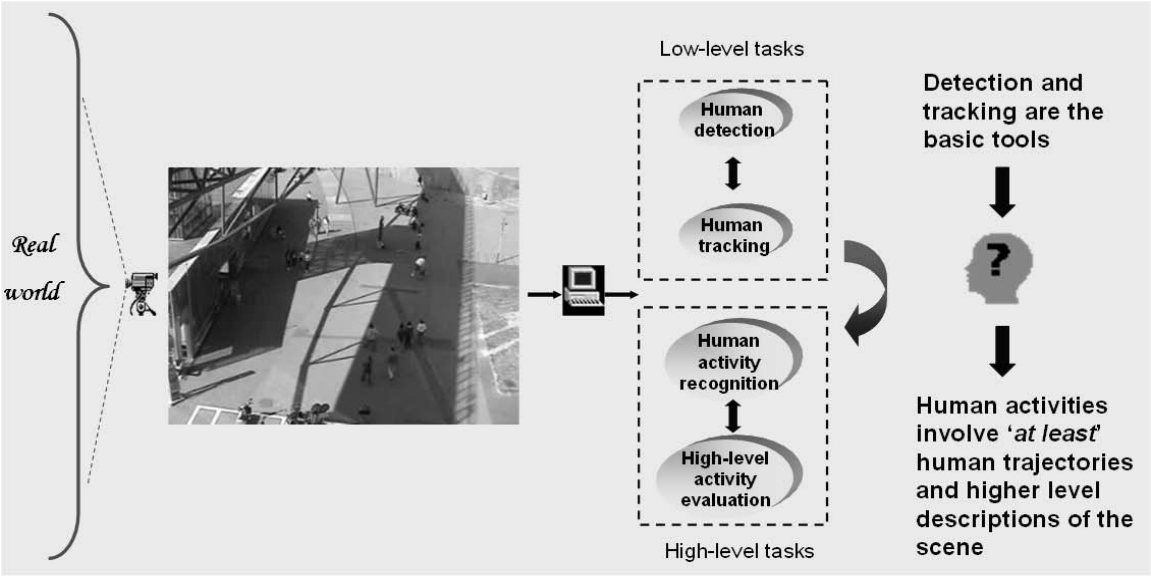


Figure 1.1: Combining low-level and high-level tasks under a common mathematical framework.

this work.

1.2 Aims and overview

The big challenge is the formulation of a mathematical framework, effective at the same time to improve the low-level tasks, namely detection and tracking of pedestrians, and which is suitable to be extended to higher level operations. In this spirit, the objective of this thesis is twofold. First, we aim to provide a behavioral model for pedestrian dynamics; second, we investigate a methodology to integrate the behavioral information with standard detection/tracking approaches.
An overview of the thesis is presented in the following. It is organized in two parts:

- Discrete choice models for pedestrian walking behavior
- Application: model-based pedestrian detection, tracking and counting

Each part is composed by different modules which are shown in Figure 1.2 and briefly described in the following sections.

1.2.1 Discrete choice models for pedestrian walking behavior

Research in the domains of crowd evacuation management, panic situation analysis, safety issues in the development of intelligent transportation systems, have triggered important issues on pedestrian dynamics and on the development of related modeling and simulation techniques. In 2001, the first international conference on Pedestrian and Evacuation Dynamics took place in Duisburg, Germany, showing the growing interest in the scientific community. A detailed revision of the state of the art on pedestrian modeling techniques is presented in Chapter 2.

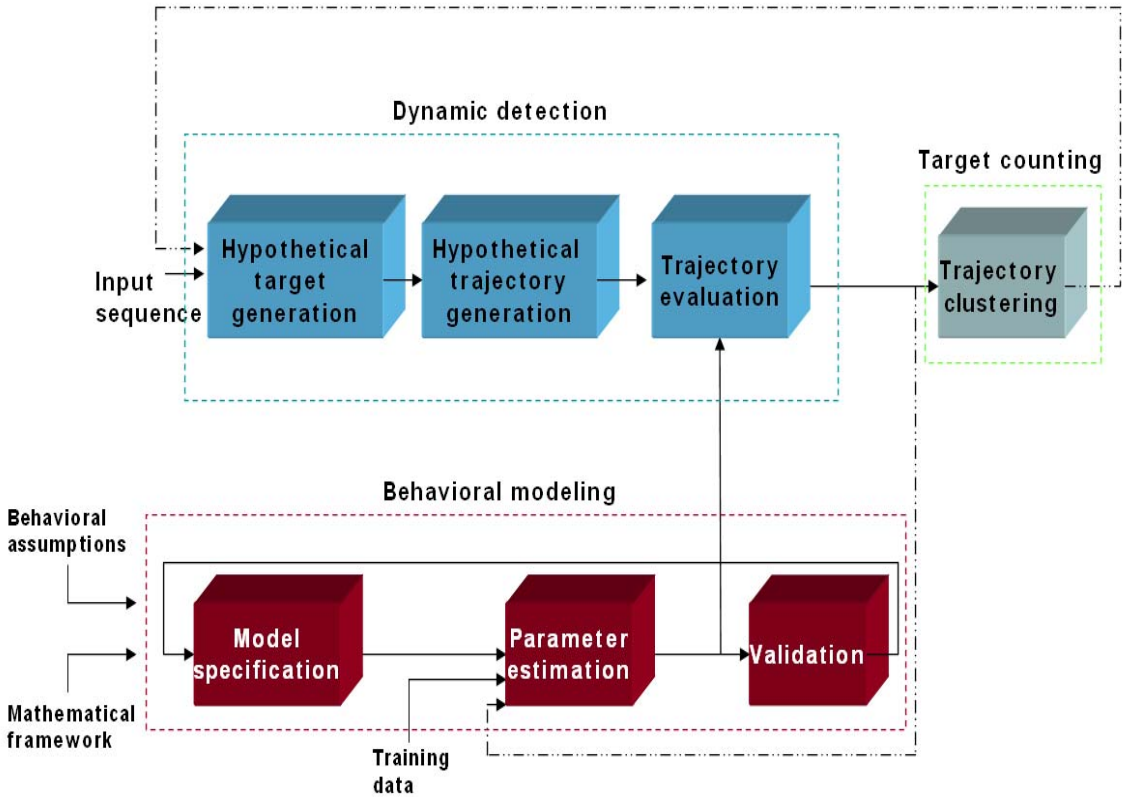


Figure 1.2: System overview. Three main modules are present (enclosed by dotted lines), behavioral modeling, dynamic detection/tracking and target counting, corresponding to the two parts. The solid lines represent the implemented links between the different parts, while the dashed lines are possible feedbacks, not yet implemented in the system. The estimated model parameters are used to evaluate the trajectories collected by the tracking itself, providing the dynamic detection output. On the other hands, the evaluated trajectories could be used for automatic data collection in new modeling procedures, while they are actually used as input data for the target counting module. Finally, the output of this module could be used to re-initialize periodically the detection/tracking system

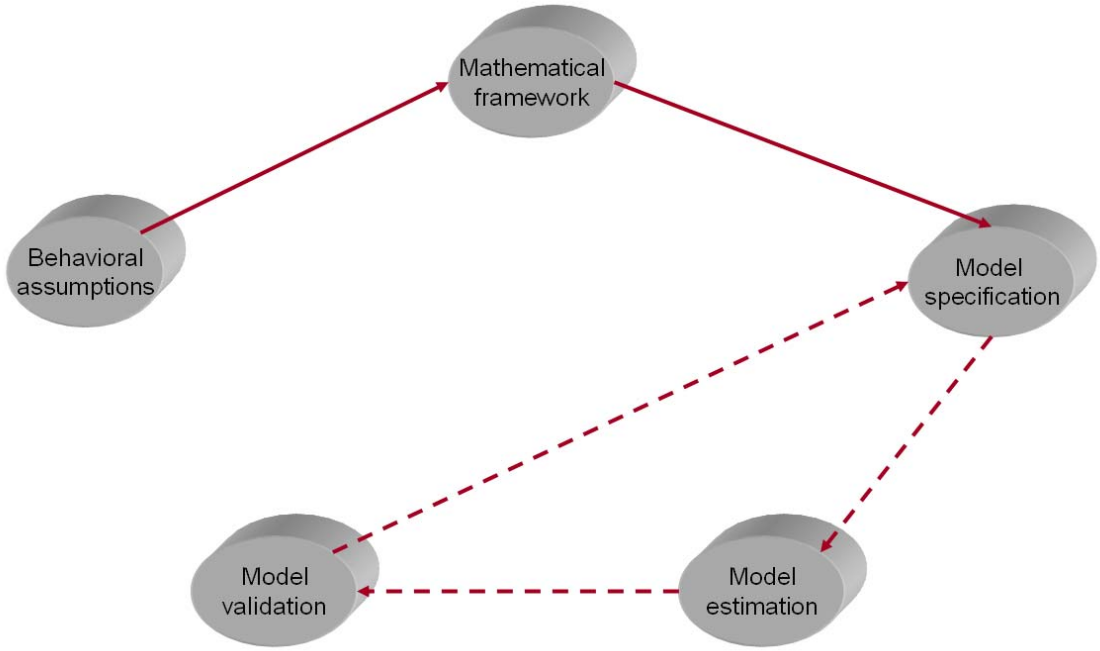


Figure 1.3: The different steps in the behavioral modeling process

Iterative process The specification and calibration of the behavioral model is the subject of Part I. It is a strongly iterative process, and it can be summarized as illustrated in Figure 1.3. As modellers, we have first to define the behavioral assumptions and the mathematical framework able to support them. We can then start with an initial model specification, including a number of unknown parameters that will have to be estimated. A final validation step is required on the coefficient estimates, in order to check their reliability. The estimated values have to be coherent with the behavioral assumptions, and consistent with the mathematical framework chosen for modeling. These last three steps, namely *specification*, *estimation* and *validation*, represent the core of the modeling process and are iterated until the final "best" solution is achieved (at least in terms of the explanatory power of the model fitting a given training dataset, and in terms of coherence with the behavioral assumptions made at the beginning).

Behavioral assumptions The walking process is considered as a sequence of choices over time, where individuals choose where they will be at the next time step, in the space around their current position. This decision making process is approached through the *rational behavior* paradigm, using Discrete Choice Models (DCM), with a Random Utility (RU) representation. Following this paradigm, individuals (decision makers) choose between different options, maximizing a certain utility function. Such a function is represented by the utility that each individual perceives from each of the available alternatives. In this context, the term *rational* is used to describe decision makers having *consistent and transitive preferences* (Ben-Akiva and Lerman, 1985). It means that repeated choices are made under identical circumstances, and the preferences over different alternatives satisfy at the transitivity property.

Mathematical framework The choice for DCM and RU models is dictated by two reasons. First, these kind of models are based on *utility maximization*, so they are consistent with the rational behavior assumption. Second, DCMs are disaggregate behavioral models, designed to forecast the behavior of individuals in choice situations, where the set of alternatives is finite. As a consequence, they well adapt to a microscopic* approach to pedestrian modeling, and they are coherent with an *agent-based* methodology, where each individual can be modeled independently. The advantages using a microscopic modeling approach will be described later on in Part I. The general theoretic framework for DCM and RU models is treated in Chapter 3.

Specification The specification of the model consists to define the utility functions, aiming to capture the short range (walking) behavior of individuals, as a response to their immediate environment. Alternative specifications are treated in Section 4.2 and Section 4.3. The use of the DCM methodology imposes the definition of the set of alternatives (*choice set*). This lead to the formulation of a *space model*. This first modeling element is adaptive with respect to individuals, meaning that each pedestrian has its own spatial representation, depending on her current speed and direction. A detailed description of the space model is given in Section 4.2. The model refers to interactions between individuals (constrained patterns) and to behaviors that are independent on the presence of other pedestrians (unconstrained patterns). The formers are captured by *leader follower* and *collision avoidance* models. The latters take into account the influence of subjective factors, independently from the other individuals. Fixed and moving obstacles are not taken into account. The reason for this "simplistic" point of view is that we have preferred a simpler but fully estimable specification, rather than more complex modeling assumptions, that did not find confirmation in the available data.

The RU formulation allows to take into account the sources of uncertainty in the modeling process, deriving by a limited knowledge of the modeller on the choice process. This lack of knowledge leads to unobserved factors, influencing the choice made step by step by walking pedestrians, choosing their spatial positions along their trajectories. This unobserved factors may give rise to correlation patterns between the different options in the choice process, and have to be taken into account. The RU approach provides us with strong econometric tools in the design of the unobserved factors influencing the behavior of decision makers. The hypotheses on the correlation structure are part of the model specification and are treated in both Section 4.2 and Section 4.3.

Estimation and validation The estimation of the unknown parameters is performed by maximum likelihood estimation (MLE), and its simulated counterpart (MSLE). Training datasets have been collected manually tracking pedestrians from video sequences, recorded in real scenarios. The collected data are described in Section 5.1 and the estimation results are presented in Section 5.2. Finally, the use of a pedestrian simulator as a validation tool for the model is described in Section 5.3.

1.2.2 Model-based pedestrian tracking

The integration of the estimated model into a tracking system is the subject of the first chapter in Part II. In the last years the problem of detection and tracking of human beings has been extensively tackled by the computer vision and image processing communities. A review of the state of the art on detection and tracking techniques is presented in Section 6.1. The tracking system is described in its parts. First, the initialization procedure to identify potential targets (pedestrians). Second,

*Microscopic approaches to pedestrian dynamics model each individual in an independent way, in contrast with the macroscopic models, where collective behaviors are taken into account, such as lanes and queues

the approach adopted to track such targets generating hypothetical trajectories is described, with particular emphasis on the methodology used and the integration of the calibrated model. Most of the existing methods for object tracking make a clear cut between object detection and tracking itself. Normally, the first task is accomplished by means of segmentation-like algorithms. This approach, which appears very natural from a conceptual point of view, has to face with the difficulties intrinsic with the segmentation problem, analyzing real scenarios. The presence of multiple targets, partial and/or total occlusions, cluttered backgrounds, along with possible low image quality and resolution, represent well known problems in detection and tracking. Often, the errors occurred and the computational time spent during this phase affect the performances of the whole system. As a consequence, we have preferred to avoid such a clear cut between detection and tracking. As indicated in Figure 1.2, after an initialization step, we track all the hypothetical targets over a certain *evaluation time*, leaving the target/non-target decision to a trajectory evaluation step. The calibrated model is used to evaluate the hypothetical trajectories, keeping only the most *human-like*. The output of this process consists in what we call *dynamic detection*, to underline the fact that pedestrians are detected on the base of their dynamic behavior over time. The shape of the tracked trajectories is so defined as a function of individual kinematic variables as well as pedestrian-pedestrian interaction parameters. The feedback from the dynamic detection module on the model parameter estimation (dashed line in Figure 1.2) illustrates how, in principle, the trajectories generated by the model-based tracker could be used as a new way to automatically collect data for new models. Numerous simplifications are made to make the integration of the model operational. The model based tracking is described into details in Section 6.2 and the relative results are shown in Section 6.3.

1.2.3 Automatic counting of pedestrians in video sequences

The subject of the second chapter in Part II is the problem of automatic counting of targets. Once we are able to track individuals with a certain accuracy, good scene descriptors such as statistical maps of the averaged pedestrian flow, density and direction could be derived. Such high level descriptors are based on the assumption to have reliable estimates of the real number of targets present in the scene. The output of a video tracking system not always provide such an accuracy, leading to a bias in the estimated number of targets. In this context, the target counting problem is fundamental for further scene analysis. The approach proposed in this third part consists in a pure post-processing of the tracked trajectories, and it is interpreted as a trajectory clustering problem. We investigate here different combinations of data representations, distance/similarity measures and grouping rules. Comparing the obtained results, interesting conclusions are drawn on this subject. The potential feedback of the counting module on the tracking system (dashed line in Figure 1.2) shows how, in principle, the clustering results could be used to re-initialize the tracker periodically.

1.3 Thesis contributions

The main objective of this research is to define a framework for pedestrian walking behavior. An application to the problem of detection and tracking is described. Finally, an analysis of the target counting problem is performed. This thesis contributes to the state-of-the-art in the following principal aspects:

- A new approach for pedestrian walking behavior modeling is proposed, based on discrete choice analysis. Several innovations are presented: an adaptive spatial discretization; a general framework including both constrained and unconstrained behaviors; specification of the leader

follower and the collision avoidance models; pameter estimation based on real pedestrian trajectory data; behavioral interpretation of the coefficient estimates.

- Application of the model in a computer vision context, namely pedestrian detection, tracking and counting in video sequences. A simple initialization scheme is proposed, avoiding complex target recognition steps; the idea of dynamic detection is presented, which consider both detection and tracking as two inter-operating tasks, in order to perform pedestrian detection using dynamic and behavioral criteria; both a deterministic and probabilistic setups for the model-based tracker are presented.

- A new approach for target counting is proposed, based on a hierarchical clustering framework. Several data representations and distance/similarity measures are compared, drawing interesting conclusions and presenting quantitative results.

Part I

Discrete Choice Models for Pedestrian Walking Behavior

Chapter 2

Pedestrian modeling: state of the art

The development of Intelligent Transportation Systems has triggered important research activities in the context of behavioral dynamics. Several new models (driving and travel behavior models), new simulators (traffic simulators, driving simulators) and new integrated systems to manage various elements of ITS, have been proposed in the past decade (see Mahmassani, 1996, Golledge, 2002). With regards to pedestrians, the focus of ITS has mainly been on safety issues (see, for example, Fuerstenberg et al., 2002), and modeling pedestrian movements in detail has rarely been considered. The ability of predicting the movements of pedestrians is valuable in many contexts. The panic situation analysis is probably the one which has motivated the large majority of research activities in the field (e.g. Helbing et al., 2000, Klüpfel et al., 2000, Helbing et al., 2002). However, it is a specific situation. Not only is the range of applications small, but also the behavior of individuals is dictated by a unique objective (save their own life) and may become irrational (Schultz, 1964, Quarantelli, 2001). Capturing the behavior of pedestrians in "normal" situations is also important in architecture (Okazaki, 1979), urban planning (Jiang, 1999), land use (Parker et al., forthcoming), marketing (Borgers and Timmermans, 1986b) or traffic operations (Nagel, in progress).
The objective of this chapter is to identify the behavioral issues arising in the context of pedestrian dynamics and analyze how they have been addressed in the literature (Bierlaire et al., 2003).

2.1 Methodological approaches for behavioral dynamics

The complexity of pedestrian behavior comes from the presence of collective behavioral patterns (as clustering, lanes and queues) evolving from the interactions among a large number of individuals. This empirical evidence leads to consider a first important distinction among the different methodological approaches: pedestrians as a flow and pedestrians as a set of individuals or agents. In the first case, the crowd is described with fluid-like properties, giving rise to **macroscopic** approaches. These models describe how density and velocity of the pedestrian flow change over time, using partial differential equations (Navier-Stokes or Boltzmann-like equations). This approach is based on some analogies observed at medium and high densities. For example, the footprints of pedestrians in snow look similar to streamlines of fluids or, again, the streams of pedestrians through standing crowds are analogous to river beds (Helbing et al., 2002). Nevertheless in these analo-

gies, the fluid-dynamic equation is difficult and not flexible. As a consequence, research focuses on the *pedestrian as a set of individuals* paradigm. This means **microscopic** models, where collective phenomena emerge from the complex interactions between many individuals (self-organizing effects).

System dynamics are based on equations describing the evolution of a system over time and have been used for both microscopic and macroscopic models. Discrete time models have the following form

$$x_{t+1} = f(x_t, \beta) \tag{2.1}$$

where x_t is a vector of state variables at time t, and β are model parameters. Continuous time models are captured by differential equations, such as

$$\frac{dx}{dt} = f(x(t), \beta) \tag{2.2}$$

where the vector of state variables $x(t)$ is a continuous function of time. This has been used for pedestrian simulation in the literature (Helbing and Molnár, 1995, Teknomo et al., 2001). In the *social forces* model of Helbing and Molnár (1995) an individual is subject to long-range forces and his dynamics follow the equation of motion, similar to Newtonian mechanics. However, we believe that complex behavioral rules and behavioral heterogeneity are difficult to capture with such models. Moreover, in most practical cases, there is no analytical solution to (2.1) or (2.2). Consequently, we have decided not to adopt this approach.

While system dynamics are time-based, **queueing models** are event-based, in the sense that they compute the state of the system for each event in a predefined agenda. The system is composed of several servers organized within a network. Each server processes items at a given rate. Items not yet processed are accumulated in queues associated with each server. The arrival of items and the service time for each server is modeled by a stochastic process. Queueing models are not appropriate to capture pedestrian dynamics. The concept of servers and the network organisation do not correspond to a tangible reality.

Game theory mimics the behavior of players, adopting a strategy knowing the strategy adopted by other players. The outcome of a game is characterized by a payoff matrix. The main focus of game theory is to identify and analyze equilibrium situations, such that no player can improve her own payoff by unilaterally changing her strategy. The approach is relevant in the context of pedestrian simulation, where the behavior in a crowd strongly depends on the behavior of other persons in the crowd. Because of the large number of "players" and the difficulty to identify an appropriate payoff matrix, we have decided to postpone the use of game theory for pedestrian simulation for future research projects.

Cellular automata allow for a time-based simulation approach, where the state of the system is represented by a regular grid composed of cells (see Toffoli and Margolus, 1987 for an introduction). Each cell can be in one of a few states (typically two, 0 or 1). At each time step, the state of each cell is updated based on its previous state and the previous state of its immediate neighbours. Therefore, it is designed for situations with local interactions. Schadschneider (2002) introduces the interesting concept of *floor field* to model the long-range forces. This field has its own dynamic (diffusion and decay), is modified by pedestrians and in turn modifies the matrix of preferences, simulating interactions between individuals and the geometry of the system. Cellular automata have been successful in the context of traffic simulation (see, for instance, Rickert et al., 1996). It is also appealing to model pedestrian behavior, and has been adopted by several authors (see,

for example, Dijkstra et al., 2000, Blue and Adler, 2001, Schadschneider et al., 2002 and Yang et al., 2002). We have decided that this approach is not appropriate in our context, due to the fixed regularity of the grid, the homogeneity of the rules and the limited number of states of each cells.

2.2 Modeling elements

Pedestrian and crowd dynamics have been studied, from an empirical point of view, for some decades using time-lapse films, photographs and direct observations. These initial efforts have yielded good empirical knowledge about the different behaviors of individuals in different kinds of environments and situations. Based on this empirical knowledge, we identify the following most important *modeling elements*.

2.2.1 Agents

An agent is an entity with its own behavior. Developed in the context of artificial intelligence (see, for instance, Ferber, 1998), agent-based models and simulation have been widely used in the context of traffic simulation (Yang and Koutsopoulos, 1997, Mahmassani et al., 1993, Barceló and Ferrer, 1997, Ben-Akiva et al., 2002). It provides a great deal of flexibility, as the behavior of each element in the system can be modeled independently, and complex interactions can be captured. All the agent-based models are also microscopic models and are based on some elementary form of intelligence for each agent (attempts to provide a *vision* and/or *cognition* capabilities). Simple behavioral rules are implemented (turning directions, obstacle avoidance) in order to reproduce more complex collective phenomena (Penn and Turner, 2002). In our context, each pedestrian is an "agent". The behavior of each agent can be modeled as a sequence of specific choices, such as the choice of the destination, the choice of the itinerary, the choice of an overall direction, or the choice of where to put the next step.

2.2.2 Space

The representation of the physical space plays a central role in the simulation. The Cellular Automata model (Schadschneider, 2002) uses a discrete structure of space. A grid of 40x40 cm^2 cells is overlapped to the area available for pedestrians, as shown in Figure 2.1. This is the typical space requirement for an individual in a dense crowd. The same grid structure is used by Kessel et al. (2002). Helbing et al. (2002), in their social force model, use the equation of motion to describe the change of location $x_i(t)$ of the pedestrian i, assuming a continuous treatment of space, similarly to the multi-layer utility maximization approach proposed by Hoogendoorn et al. (2002). In all these models the pedestrian is seen as a point or a particle in a 2D environment. With the recent development in rendering techniques and Virtual Reality simulations, other models are based on a 3D representation. In the agent-based approaches, the agent moves through a virtual environment where the movements can be discrete or continuous (Thalmann and Bandi, 1998, Penn and Turner, 2002).

A completely different approach is proposed by Borgers and Timmermans (1986b). They use a network representation (Figure 2.2), where each node corresponds to a city-center entry point or a departure point and each link denotes a different shopping street. In this case, the network topology represents the walk-able space and any movement occurs along the links between two consecutive nodes.

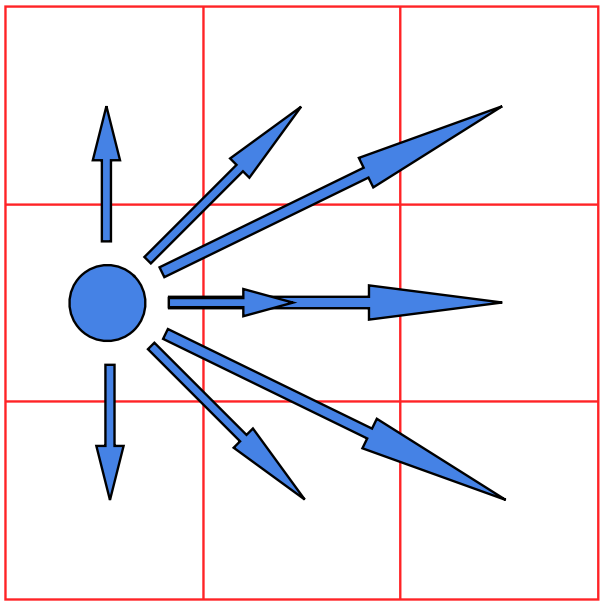


Figure 2.1: Example of a static grid discretization

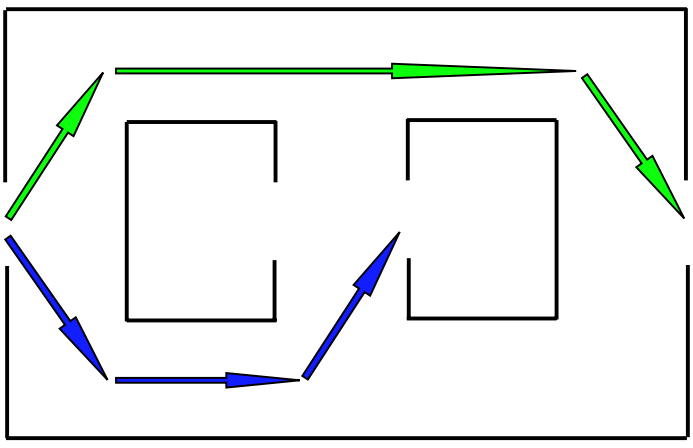


Figure 2.2: Example of a network-based spatial representation

The space model is directly connected with the concept of visual field. Indeed, pedestrians are influenced by what they actually see. More precisely, the complex cognitive mapping and learning process in a human being, find their first source of inputs from the images of the space around. Some attempts of explicit modeling have been attempted in the robotics community (Bachelder and Waxman, 1994), who simulate the hippocampal cognition process with a really high computational cost. Important studies on the interactions between pedestrians and the space are made by Timmermans group (Essyx tool), while a pedestrian simulator that tries to reproduce the visual field is EVAS (Penn and Turner, 2002). The EVAS approach is based on a development of the *Space Syntax* theory (Hillier and Hanson, 1984 and Hillier et al., 1993): the Visibility Graph Analysis (VGA). The configuration of the space seems to be one of the first sources in the variance of human behavioral patterns. The VGA states that, given a grid of points that covers the space layout, it is possible to create a graph, the *visibility graph*. Each grid's point represents a node and two nodes are connected if and only if they are mutually visible (another possibility is *one-depth* visibility that is, visible by means of an intermediate node). From the graph, it is possible to calculate some features of the space that are important for pedestrian way-finding. The first of these coefficients is the *Neighbourhood size* N_i defined as the set of directly visible vertices:

$$N_i = \{v_j : e_{ij} \in E\} \tag{2.3}$$

where N_i is the neighbourhood of location v_i, e_{ij} is the link between nodes v_i and v_j and E is the set of all links in the graph. The neighbourhood size value is proportional to the visible area and its plot draws contours of equal viewable areas across the space (Turner et al., 2001). Another important measure is the *clustering coefficient* C_i for the neighbourhood N_i (of size k_i) of location v_i:

$$C_i = \frac{|\{e_{jk} : v_j, v_k \in N_i \wedge e_{jk} \in E\}|}{k_i(k_i - 1)} \tag{2.4}$$

It is the number of edges between all the vertices in the neighbourhood of vertex v_i divided by the total number of possible connections with that neighbourhood size. It can be seen also as a measure of the inter-visible space within the visibility neighbourhood of a point. Again if we think about a pedestrian in position v_i, the C_i values give a measure of the potential to form groups or to interact (Turner et al., 2001). Conroy (2001) have further shown that in multi- directional visual field areas, as junctions, there is a strong correspondence with the stopping behavior of people (thinking about a new direction). Many other features can be extracted from the graph, as the mean shortest path length, giving to the pedestrian a crude form of cognition or memory.

Obstacles and attractors must be also modeled. Fixed obstacles are represented by regions that no pedestrian can access. Moving obstacles are (groups of) other pedestrians occupying some space which is consequently not available anymore. Attractors are useful areas with particular meaning for the individuals. Examples of attractors can be the shopping windows or areas becoming interesting because of the presence of painters or musicians. Helbing et al. (2002), in their social force model, take into account time-dependent attractive interactions. They use social forces with an interaction range longer and a strength parameter smaller compared with repulsive interactions. So, practically, they define attractors as long-range forces.

Finally, origin and destination areas, where pedestrians enter and exit the system must be defined. Those could be doors, elevators, stairs, and of course the boundaries of the modeled area.

2.3 Behavior

Several decisions can be made by pedestrians. Each of them can be associated to a different behavioral model.

2.3.1 Destination choice

The destination choice problem is tricky in the context of pedestrian simulation. Indeed, some individuals may not have a destination at all if, for instance, they are walking around waiting for a bus. In shopping areas, the destination may change rapidly depending on the environment or on the attractors (see Whynes et al., 1996, Dellaert et al., 1998).

Borgers and Timmermans (1986b) propose a simulation of pedestrians in the shopping streets of the city centres. The model is a Monte Carlo simulation which implies that the behavior of each individual is simulated by a series of draws of random numbers from successive probability distributions. In their work, the authors build different sub-models related to the number of goods bought by pedestrians, in which retail sector, in which link of the urban network (the link is the shopping street). As an example, we report the link-choice model:

$$p_{nl}^{g} = \frac{(\sum_{m \in l} F_m^g)^{\alpha} \exp\left(-\beta \min\left[\sum_{l'' \in r} d_{l''}\right]\right)}{\sum_{l'=1}^{L} \{(\sum_{m \in l'} F_m^g)^{\alpha} \exp\left(-\beta \min[\sum_{l'' \in r} d_{l''}]\right)\}} \tag{2.5}$$

where p_{nl}^{g} is the probability that a good in retail sector g will be bought at link l providing that the pedestrian departed from city entry point n, F_m^g is the total amount of floor-space in retail sector g at destination $m(m = 1, 2, ..., M)$, $\min\left[\sum_{l'' \in r} d_{l''}\right]$ is the distance associated with the shortest route from city entry point n to link l, and α,β are parameters to be estimated.

Kopp (1999) uses in the EVACSIM simulator the so called *Primary/Secondary* destination selection. A shortest path algorithm, using a sub-goal system, was developed for this simulator to allow people to effectively navigate around obstacles. If a person's path to an exit destination is blocked (checked with a line intersection test), the person finds a sub-goal that is in a line-of-sight with the person. If multiple sub-goals are in a line-of-sight, the person chooses the one that will lead to the shortest path to the exit destination. This approach to model the *destination selection* is local and captures an obstacle avoidance behavioral pattern. It does not deal with any "high level" decision process as for example trip purpose or activity-based scheme. Moreover, EVACSIM is mainly oriented to evacuation situations, where the pedestrians have one or multiple exits and their behavior is essentially an event response pattern (for example fire in a building).

2.3.2 Route Choice

Borgers and Timmermans (1986a) addresses the route choice problem as an utility maximization problem. The *objective characteristics* X_{lk} of the link l are transformed into subjective perceptions or evaluations by means of a functional relationship f_k:

$$x_{lk} = f_k\{X_{lk}\}, k = 1, 2, ..., K. \tag{2.6}$$

After that, the *subjective utility* $U(l)$ is obtained as an algebraic combination of the subjective values:

$$U(l) = h(x_{lk}), k = 1, 2, ..., K. \tag{2.7}$$

Likewise, the route's utility equals:

$$U(r) = h'(U(l); d_r), l \in r \tag{2.8}$$

where h' is another algebraic function and d_r is the total subjective distance associated with route r. The pedestrian will choose the route that will maximizes his subjective utility.

Blue and Adler (2001) analyse the problem from a *self-organizing* point of view. They use a CA model, with a limited rule-set for the pedestrian behavior and look at the emergent collective behaviors. Their route-choice is lane-based. They show how unidirectional, bi-directional, cross-directional and 4-directional pedestrian flows emerge from CA simulations. The model centres on a two-stage parallel update process whereby lane assignment and forward motions change the positions of all pedestrians in two parallel update stages in each time step. They assume that only pedestrians in the immediate neighbourhood affect the movement of a pedestrian. We believe that this assumption is valid only at the operational level in a hierarchical model (the local interactions among pedestrians).

There is a distinction between the individuals who know their destinations and the others who do not have a precise destination. All the models seen until now, refer to the first category. We could talk about *explorers* referring to people who do not have a specific destination. It is clear that, in this case, the behavioral patterns are different and other parameters become important. Penn and Turner (2002), in their EVAS simulator (agent-based), address this kind of population. The agent takes a decision about her destination and chooses the route every three steps, basing the decision process on the simulated visual field. The direction is chosen randomly inside the visual field. We think that this is a limitation and is applicable to the only exploratory pattern. It would be interesting to apply these concepts in a discrete choice model framework, using the visual field as a source of exogenous information and mixing it with endogenous parameters.

We conclude this section saying that the idea to address the destination and/or route choice problems in a pedestrian behavior context, stems from previous research activities, namely in the Intelligent Transportation System context. Among the route choice literature, we refer the reader to Ben-Akiva et al. (1984), Charlesworth and Gunawan (1987), Bovy and Stern (1990), Cascetta et al. (1992), Ben-Akiva and Bierlaire (1999) and Ramming (2001). As already mentioned in the introduction, several new models capturing driver behavior and traveler behavior, as well as traffic simulators, have been extended to the pedestrian behavior and way-finding problems (Muramatsu et al., 1999).

2.3.3 Speed

In an empty space, the destination and the path are almost sufficient to reproduce the trajectory of a given pedestrian. Almost, because an average speed must be assumed, and small random deviations from the given path must be allowed for the sake of realism. When the environment is crowed and contains obstacles, the direction and the speed of the pedestrian may be significantly affected.

In most models in the literature, the important parameters influencing the behavior are the *desired speed* and speed regimes. In the social force model, the desire to adapt the actual velocity $v_i(t)$ to the desired speed v_t^0 within a certain "relaxation time" τ_i is reflected by the *acceleration term* $[v_i^0 e_i^0 - v_i(t)]/\tau_i$. The contribution $v_i^0 e_i^0/\tau_i$ is interpreted as a *driving term* and $-v_i(t)/\tau_i$ as a *friction term*. The CA model addresses the problem using the dynamic of the *floor field*. So, the movement of a pedestrian is considered as the movement of a particle that crosses a field with its own dynamic (diffusion and decay). In the lane-based approach, Blue and Adler (2002) design the model to

account for variations in walking speeds observed in the real world. Each time step is one second and they consider walking speed varying among pedestrians, using distribution of walking speed with a cell size of 0.21m^2:

1. *fast walkers*: maximum speed of 4 cells per time step (about 1.8m per time step);
2. *standard walkers*: maximum of 3 cells per time step (1.3m per time step);
3. *slow walkers*: maximum of 2 cells per time step (0.85m per time step).

In their experiments, they use a population composed of 5% of fast, 90% of standard and 5% of slow walkers. In the multi-layer utility maximization model, Hoogendoorn et al. (2002) define the kinematics of the pedestrian as follows:

$$dx = vdt + \sigma dw \tag{2.9}$$

where $v = v(\tau)$ is the velocity vector for $\tau > t$. The term w is a Wiener process and denotes the *uncertainty in the expected traffic conditions* and its effects on the pedestrian's kinematics. The speed of the individual is limited by the physical conditions and by the other individuals. There is a set of *admissible velocities* defined as

$$V_a(t,x) = \{v : \|v\| \leq v_0(t,x)\} \subset \mathbb{R}^2 \tag{2.10}$$

The t and x dependence describes the changing of the maximum speed stemming from the change in flow conditions as well as differences in maximum speed between different parts of the walking infrastructure. Last, but not least, the maximum speed of a specific pedestrian also depends on the individual's characteristics (age, gender, trip-purposes, luggage etc...).

2.3.4 Pedestrian interactions

Collision avoidance The *collision avoidance* pattern stems automatically from a combination of the velocity vector of the other pedestrians and the density parameter. In microscopic models, an individual tries to keep a minimum distance from the others ("territorial effect"). In the social force model, this pattern is described by repulsive social forces:

$$\mathbf{f}_{ij} = A_i \exp\left[(r_{ij} - d_{ij})/B_i\right]\mathbf{n}_{ij}(\lambda_i + (1-\lambda_i)\frac{1+\cos\varphi_{ij}}{2}) \tag{2.11}$$

where A_i is the interaction strength, B_i the range of the repulsive interaction, d_{ij} the distance between pedestrians i and j, r_{ij} the sum of the radii, $\mathbf{n}_{ij}$ the vector pointing from i to j, the angle φ_{ij} denotes the angle between the direction of motion and the direction of the object exerting the repulsive force. Finally, the parameter λ_i takes into account the fact that the situation in front of a pedestrian has a larger impact than things happening behind.

Crowd effects One of the first approaches to describe the crowd effect was in Reynolds (1987), the *leader-follower* pattern. One or more individuals follow another moving individual designated as the leader. Generally the followers want to stay near the leader, without crowding the leader, and taking care to stay out of the leader's way (in case they happen to find themselves in front of the leader). In addition, if there is more than one follower, they want to avoid bumping into each other. The implementation of leader following relies on *arrival behavior*, a desire to move toward a point, slowing down as it draws near. The arrival target is a point offset slightly behind the leader

(the offset distance may optionally increase with speed). If a follower finds herself in a rectangular region in front of the leader, she will steer laterally away from the leader's path before resuming arrival behavior.

In the social force model of Helbing et al. (2002), physical interaction forces come into play when pedestrians get so close to each other that they have physical contact. This is mainly the case in panic situations but also as a reaction to an event. The authors assume a *body force* $k(r_{ij} - d_{ij})\mathbf{n}_{ij}$ counteracting body compression and a *sliding friction force* $k(r_{ij} - d_{ij})\Delta v^t_{ij}\mathbf{t}_{ij}$ impeding relative tangential motion. The model for this effect comes from the granular interaction formula:

$$\mathbf{f}^{ph}_{ij}(t) = k\phi(r_{ij} - d_{ij})\mathbf{n}_{ij} + k_1\phi(r_{ij} - d_{ij})\Delta v^t_{ij}\mathbf{t}_{ij} \tag{2.12}$$

where the function $\phi(x)$ is equal to x if $x \geq 0$ and 0 otherwise. The vector $\mathbf{t}_{ij}$ is the tangential direction and Δv^t_{ij} is the tangential velocity difference. The r_{ij} term is the sum of the radii r_i and r_j, d_{ij} is the distance between the centres of mass of pedestrians i and j and k,k_1 are some large constants.

A large amount of literature in transportation science concerning *car following* models has inspired our analysis of the leader follower behavior for pedestrians. The main idea in these models is that two vehicles are involved in a car following situation when a subject vehicle follows a leader, normally represented by the vehicle in front, reacting to its actions. Normally, a sensitivity-stimulus framework is adopted. According to this framework a driver reacts to stimuli from the environment, where the stimulus is normally chosen as the leader relative speed. Different models differ in the specification of the sensitivity term (see Ahmed, 1999, Herman and Rothery, 1965, Lee, 1966, Newell, 1961 and Toledo, 2003 among the others). The sensitivity-stimulus approach models acceleration behaviors. In the case of cars, it is natural a distinction between acceleration and direction change (lane change) behaviors, being imposed by the transport facility itself. On the other hand, the pedestrian case is more complex being the movements purely two-dimensional movements on the walking plane, where accelerations and direction changes are not easily separable.

Interesting studies on human interactions have been conducted in human sciences and psychology, providing a sociological interpretation of crowd effects. A related and interesting concept is that of *personal space* (see Horowitz et al., 1964, Dosey and Meisels, 1969 and Sommer, 1969). Personal space is a protective mechanism founded on the ability of the individual to perceive signals from one's physical and social environment. Its function is to create the spacing patterns that regulate distances between individuals and on which individual behaviors are based (Webb and Weber, 2003). Several studies in psychology and sociology show how individual characteristics influence the perception of the space and interpersonal distance. Brady and Walker (1978) found for example that anxiety states are positively correlated with interpersonal distance. Similarly, Dosey and Meisels (1969) found that individuals establish greater distances in high-stress conditions. Hartnett et al. (1974) found that male and female individuals approached short individuals more closely than a tall individual. Other studies (Phillips, 1979 and Sanders, 1976) indicate that the other person's body size influences space. Hence, it seems natural that individual characteristics as age, sex, weight among others influence the spatial perception, interpersonal distance and human-human interactions. Unfortunately, such a kind of information is very hard to collect, requiring the setting of controlled experimental conditions.

2.3.5 Hierarchical decision process for pedestrians

A very important work, especially from a methodological point of view, is that proposed by Hoogendoorn et al. (2002), Hoogendoorn (2003) and Daamen (2004). The authors proposed a very plausible framework for the decision process related to walking pedestrians. They assume that people make decisions at three different levels. The highest level is called *strategical*, where people decide about the activities they want to perform. Some of those activities are discretionary, some other are mandatory (Dijkstra et al., 2000, Arentze and Timmermans, 2004, Penn, 2003). The second level is called *tactical*. Here people, given the activities chosen at the strategical layer, decide about the order of the activity execution (the *activity scheduling* process); they can also decide about where to perform the different tasks involved in their planned activities (*activity area choice* process). These first two decision making processes in the tactical level are actually more conceptual than operational. Few literature exists about these topics and, more important, few models have been calibrated on real data. Also for the objective difficulties to collect related data, often such a choices are assumed to be exogenous in the models. The most important tactical level process is the *route choice* between origins and (intermediate) destinations. The principle is the utility maximization (more precisely the *expected disutility minimization*) taking into account different route attributes such as travel time, distance travelled, safety, comfort, etc.... They do not use a discrete choice framework; the number of choice options is infinite in continuous time and space. We show, as an example, the equation used to describe the expected cost C_i:

$$C_i(T_i, v_{[t,T_i)}|T_{i-1}, x(T_{i-1})) = E\left[\int_t^{T_i} L(\tau, x(\tau), v(\tau))\, d\tau + \phi(T_i, x(T_i))\right] \tag{2.13}$$

The time interval $[t, T_i)$ denotes the interval between the current time and the arrival time at an activity area while $v[t, T_i)$ is the velocity path. The *running cost* $L(\tau, x(\tau), v(\tau))$ shows the costs incurred during the time interval $[\tau, \tau + d\tau)$ where $x(\tau)$ is the location and $v(\tau)$ is the applied velocity to change the position. The *terminal cost* $\phi(T_i, x(T_i))$ reflects the cost due to ending up at position $x(T_i)$ at the terminal time T_i. The terminal cost represents the penalty ϕ_i for not having arrived in time at any of the activity areas. The running cost is related to the different contributions of the different route attributes (discomfort, walking at certain speed, expected number of pedestrian interactions etc...).

The last level in the conceptual hierarchy is the *operational* one. People make instantaneous decisions for the next time step, given the choices made at the two higher levels. Individual walking behavior is defined here, where pedestrians are involved in managing interactions with other people, with the surrounding environment (fixed and moving obstacles, depending on the context), waiting at the activity areas and performing activities themselves.

2.4 Simulation

We report here some previous works on pedestrian simulation. Several programs have been developed to simulate pedestrian behavior in various contexts, but the great majority of them seems to be aimed at building evacuation, especially in case of fire. Given that a full discussion on pedestrian simulation techniques is out of the scope of this thesis, we provide here only some references to some of the existing simulation systems, referring the interested reader to the reported references:

- PEDSIM is a software tool for microscopic pedestrian and crowd simulation. It implements the social force model described in Helbing and Molnár (1995).

- NOMAD is a microscopic simulator developed at the Transport & Planning department of the Delft University of Technology. It is based on microscopic characteristics of pedestrians, such as walking speed, pedestrian size, etc....
- EXODUS (Gwynne et al., 1997) is a software tool for simulating the evacuation of large numbers of people from buildings, airplanes, boats, etc.
- Simulex (Thompson and Marchant, 1994) is the evacuation simulation part of the IES Virtual Environment, a set of tools developed to aid in the design and evaluation of buildings.
- EVAS (Turner and Penn, 2002) is a pedestrian simulator based on Visibility Graph Analysis.
- EVACSIM (Kopp, 1999) is an evacuation oriented simulator developed in Java. It is freely available, including source code.

Pedestrian modeling and simulation is becoming an important field of research, as the range of possible applications is widening. In this chapter, we have tried to identify the behavioral aspects of pedestrian dynamics, describing how they have been addressed in the existing literature. Some analogies with traffic modeling and simulation techniques arise from this state-of-the-art review. Problems such as destination choice or route choice are well known to the transportation science community.

Being inspired by traffic modeling techniques, our approach interprets the short range behavior of pedestrians as a sequence of choices, in a discretized spatial representation. As already mentioned in Chapter 1, this idea is naturally implemented by means of a DCM approach.

Chapter 3

Discrete Choice Models and Random Utility Models

In this chapter we introduce discrete choice models (DCM) and random utility models (RUM), giving an overview on the theoretical aspects and focusing on those models that will be used in the following chapters to model the pedestrian walking behavior. The use of DCM in our context is justified by several reasons. First of all, we model the walking process as a sequence of choices on *where to put the next step*, so DCMs represent a natural way to deal with this modeling assumption. Moreover, they are disaggregate models, hence well adapted to a microscopic approach for pedestrian behavior, and coherent with an *agent-based* methodology, where each individual can be modeled independently. Finally, being econometric models, they are designed to be calibrated on real data, providing a set of statistical techniques to model the uncertainty. DCMs have been widely used in econometrics (McFadden, 1978; Manski and McFadden, 1981; Koning, 1991; Koning and Ridder, 1994; Hensher and Johnson, 1981) with numerous applications to different fields, such as marketing, finance and labour economics. Large use of DCMs has also been done in transportation science (Ben-Akiva and Lerman, 1985; Ben-Akiva and Bierlaire, 1999; Ben-Akiva et al., 1984; Cascetta et al., 1992) over the last three decades and more (route choice, destination choice, departure time choice). In 2000, Daniel McFadden won the Nobel prize in Economics for his contributions on discrete choice analysis, increasing furthermore the interest of the scientific community on this modeling techniques.

The chapter is structured as follows. Section 3.1 and 3.2 give an overview on the most important behavioral assumptions for DCM, while the remaining sections describe the different model specifications.

3.1 Individual choice behavior

The interest here is focussed on individual choice behavior, i.e. a *disaggregate* approach is adopted. Such a choice behavior theory is (1)*descriptive*, postulating how individuals behave and not how they should behave, (2)*abstract*, i.e. generalisable and (3) *operational*, resulting in a certain number of parameters estimated from data (Ben-Akiva and Lerman, 1985). The main assumptions about individual choice behavior are:

- decision maker, including her socio-economic characteristics,

- the set of alternatives (choice set),
- attributes of alternatives,
- decision rules.

The decision maker is the unit of decision and can represent a single individual, a household, a firm, an organization. Decision makers face to different choices and have different tastes. The choice set represents the set of alternatives which are known to the decision maker and are available during the decision process. The attributes describe the alternatives in terms of their attractiveness. Finally, the decision rule describes the mechanisms internal to the decision maker that are used to process the available information, arriving to a unique choice.

3.2 Rational behavior and random utility

Despite other existing approaches, the decision rule here is assumed to be deterministic. More precisely, individuals choose among the alternatives, maximizing the *utility* they perceive from each of them. This index captures the attractiveness for an alternative, and can be reduced to a scalar value. It is based on the notion of trade-offs used by the decision maker comparing different attributes. This behavioral paradigm is called *rational*, meaning that the decision maker shows consistent and transitive preferences. Consistent in the sense that she will repeat the same choice when faced to identical situations; transitive in the sense that if in the choice set alternative one is preferred to alternative two, and alternative two is preferred to alternative three, then alternative one is also preferred to alternative three. Nonetheless, some inconsistencies are observed in choice experiments. They are assumed to be the result of a lack of knowledge on the part of the analyst. While the individual is always assumed to choose the alternative with the highest utility, modellers do not know with certainty the utility values, which are then treated as random variables. This approach has been formalized by Manski (Manski, 1977), identifying four sources of randomness: unobserved alternative attributes, unobserved socio-economic characteristics, measurements errors and instrumental variables.

3.3 Random utility models

Given a population of N individuals, making choices between a choice set C_n with J alternatives, we define the (random) utility function U_{in} perceived by individual n for alternative i as follows:

$$U_{in} = V_{in} + \varepsilon_{in} \tag{3.1}$$

with $i = 1, ..., J$ and $n = 1, ..., N$. V_{in} represents the deterministic part of the utility and it is a function of the alternatives' attributes and the socio-economic characteristics of the decision maker. Normally, the term V_{in} is modeled with a linear-in-parameters specification. The ε_{in} term is a random variable capturing the uncertainty. Under the utility maximization assumption, the output of the model is represented by the choice probability that individual n will choose alternative i, given the choice set C_n. It is given by:

$$\begin{aligned} P_n(i|C_n) &= P_n(U_{in} \geq U_{jn}, \forall j \in C_n) \\ &= P_n(V_{in} + \varepsilon_{in} \geq V_{jn} + \varepsilon_{jn}, \forall j \in C_n) \\ &= P_n(\varepsilon_{jn} - \varepsilon_{in} \leq V_{in} - V_{jn}, \forall j \in C_n) \end{aligned} \tag{3.2}$$

Looking at Equation 3.2, we can conclude that in order to define the choice probability, only the difference between the utilities matters. Moreover, multiplying the utilities by a positive constant does not affect the choice probability. Different DCMs are obtained making different assumptions on the error terms.

3.4 The Multinomial Logit model

The Multinomial Logit model (MNL) is largely the simplest and most used discrete choice model. The MNL is obtained by assuming that each ε_{in} in the utility functions is independently and identically Gumbel distributed. This distribution, also called type I extreme value distribution, is characterized by the density

$$f(\varepsilon_{in}) = \mu e^{-\mu(\varepsilon_{in}-\eta)} e^{-e^{-\mu(\varepsilon_{in}-\eta)}} \tag{3.3}$$

where η is a location parameter and μ is a positive scale parameter. Under these assumptions, the choice probability is given by the following expression

$$P_n(i|C_n) = \frac{e^{\mu V_{in}}}{\sum_{j \in C_n} e^{\mu V_{jn}}} \tag{3.4}$$

with $j \in C_n$ and assuming a null location parameter (which is not a restrictive assumption). The MNL model provides a closed form solution for the choice probabilities, which is a clear advantage against other models, as for example the Probit model (see Ben-Akiva and Lerman, 1985 and Train, 2003). However, despite its large use in the literature, it shows some limitations that can be severe in certain situations.

MNL limitations For the purposes of our work, we describe here three important characteristics of the MNL, which limit its flexibility and induce the use of more sophisticated techniques. They are:

- independence from irrelevant alternatives (IIA);
- deterministic taste variations;
- homoscedasticity.

The IIA property derives from the independence assumption on the error terms. It can be formulated as follows: *the ratio of the choice probabilities for two alternatives is not affected by the systematic utilities of the other alternatives*. This property is the most important limitation of the MNL model. As modellers, we do not observe all the attributes for each alternative. It is possible that two or more alternatives, sharing common unobserved attributes, give rise to correlation patterns in the choice set, violating the IIA assumption.
The second limitation concerns with *taste variations*. It seems plausible that the behavior in choice

situations varies across the population, and it is fundamental for a model to capture such variations. The MNL model can capture only deterministic taste variations, occurring when for example the population can be deterministically divided into different segments. This is not always the case, especially when relevant data are not available for all the segments.
The third limitation, i.e. the homoscedasticity, is imposed by the assumption that the error terms are identically distributed. It means that all of them have the same scale parameter μ. This fact implies that the variances are the same across the population. This assumption represents a limitation in those cases, for example, where different sources of data are combined. This is actually the problem in our case. We will use two different sets of pedestrian trajectories, collected separately in Switzerland and Japan at different time, and pooled together in an unique dataset, as explained in Chapter 5.1.

In the following, Section 3.5 and Section 3.6 describe different models which allow to overcome the first limitation of the MNL model. In Section 3.7 we deal with the second limitation, while the third one is addressed in Section 3.8.

3.5 Generalized Extreme Value models

Common unobserved attributes shared between alternatives give rise to correlation patterns in the choice set. The Generalized Extreme Value model family (GEV) allows to take into account these patterns. These models provide a closed form solution for the choice probabilities, allowing at the same time for flexible correlation structures. GEV models have been introduced by McFadden (1978) and are characterized by a multivariate extreme value distribution (Bierlaire, 2005). The distribution of the vector ε of the error terms is characterized by the following cumulative distribution function:

$$F_\varepsilon(V_1, ..., V_J) = e^{-G(e^{-V_1}, ..., e^{-V_J})} \tag{3.5}$$

where G is a differentiable function defined on R_+^J with the following properties:

- $G(y) \geq 0$ for all $y \in R_+^J$;
- G is homogeneous of degree $\mu > 0$, that is $G(\alpha y) = \alpha^\mu G(y)$;
- $lim_{y_i \to +\infty} G(y_1, ..., y_i, ..., y_J) = +\infty$, for each $i = 1, ..., J$;
- the kth partial derivative with respect to k distinct y_i is non-negative if k is odd and non-positive if k is even, i.e. for any distinct indices $i_1, ..., i_k \in 1, ..., J$, we have

$$(-1)^k \frac{\partial^k G}{\partial y_{i_1} ... \partial y_{i_k}}(x) \leq 0, \forall y \in R_+^J \tag{3.6}$$

where $y_i = e^{V_i}$. The choice probability for alternative i to be chosen by a given decision maker is:

$$P(i|C) = \frac{y_i G_i(y_1, ..., y_J)}{\mu G(y_1, ..., y_J)} \tag{3.7}$$

Several models can be derived from this general formulation, through an opportune specification of the generating function G. The MNL model, the Nested Logit (NL) and the Cross Nested (CNL) logit models can all be derived from the GEV formulation. We describe here only the CNL because it will be used in the following chapters. For more details on the derivation of the other models we remind the interested reader to Ben-Akiva and Lerman (1985), Train (2003) and Wen and Koppelman (2001).

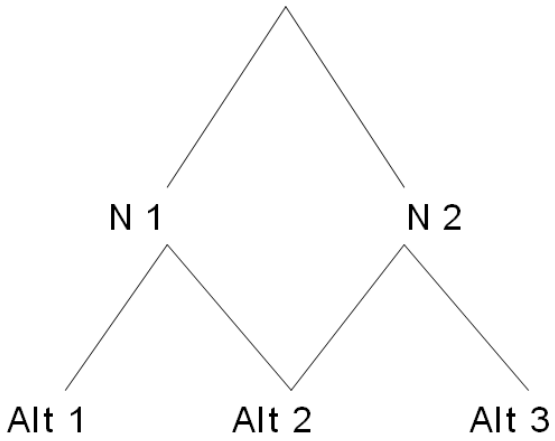


Figure 3.1: An example of cross nested logit model. Alternative 1 belongs to nest 1 and alternative 3 belongs to nest 2. Alternative 2 belongs to both nests 1 and 2, and it is correlated with both alternatives 1 and 3.

3.5.1 The Cross Nested Logit model

The CNL model is a GEV model based on the following generating function (Abbé et al., 2005)

$$G(y_1, ..., y_J) = \sum_{m=1}^{M} \left(\sum_{j \in C} (\alpha_{jm}^{1/\mu} y_j)^{\mu_m} \right)^{\frac{\mu}{\mu_m}} \tag{3.8}$$

where $\alpha_{jm} \geq 0, \forall j, m$, $\sum_{m=1}^{M} \alpha_{jm} > 0, \forall j$, $\mu > 0$, $\mu_m > 0, \forall m$ and $\mu \leq \mu_m, \forall m$. This formulation leads to the following expression for the choice probability, using $y_i = e^{V_i}$:

$$P(i|C) = \sum_{m=1}^{M} \frac{\left(\sum_{j \in C} \alpha_{jm}^{\mu_m/\mu} y_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}}}{\sum_{n=1}^{M} \left(\sum_{j \in C} \alpha_{jn}^{\mu_n/\mu} y_j^{\mu_n} \right)^{\frac{\mu}{\mu_n}}} \frac{\alpha_{im}^{\mu_m/\mu} y_i^{\mu_m}}{\sum_{j \in C} \alpha_{jm}^{\mu_m/\mu} y_j^{\mu_m}} \tag{3.9}$$

For each j and m, α_{jm} represents the degree of membership of alternative j to nest m. The nested logit model is a special case, where $\alpha_{jm} = 1$ if alternative j belongs to nest m, and 0 otherwise. In the CNL formulation flexible correlation patterns are allowed among the alternatives. An example is shown in Figure 3.1. In this case we have three alternatives and two nests. The second alternative belongs to both nests, and it is correlated with both the first and third alternatives. This correlation structure would not be allowed, for example, with a simpler nested logit model, where each alternative can belong to only one nest.

3.6 The Logit Kernel Error Component model

The GEV models allow to relax the IIA property of the MNL model, still keeping a closed form solution for the choice probabilities. However, in order to reproduce more general correlation structures between the alternatives, more flexible models have been developed. The Logit Kernel model family attains this goal. While several specifications have been presented in literature

(see Walker, 2001; Ben-Akiva and Bolduc, 1996; McFadden and Train, 2000; Train, 2003; Hensher and Greene, 2001), the one we describe here is the *error component* specification. In this model the vector of the error terms ε_n for an individual n is specified as follows:

$$\varepsilon_n = F_n \xi_n + \nu_n \tag{3.10}$$

where ξ_n is an $(M \times 1)$ vector of M multivariate latent factors, F_n is a $(J_n \times M)$ matrix of fixed factor loadings equal to 0 or 1 and ν_n is a $J_n \times 1$ vector if i.i.d Gumbel terms, with J_n the number of alternatives. Generally, we assume the factors as independent normally distributed, such as

$$\xi_n = T\zeta_n \tag{3.11}$$

where ζ_n are a set of standard independent normal terms, and TT' is the covariance matrix of ξ_n (being T the Cholesky factorization). In Equation 3.10, the error term is made up of two components: the first one has a multivariate distribution, capturing the interdependencies between the alternatives, and the second is the standard Gumbel term of the logit model. This specification is interesting because conditioning on the ζ_n, the choice probability becomes equal to a MNL model (i.e., logit kernel). The unconditional choice probability is then obtained through integration of the conditional one, over all the possible values of the multivariate ζ:

$$P(i) = \int_{\zeta} \Lambda(i|\zeta_n) n(\zeta, I_M) d\zeta \tag{3.12}$$

where with $\Lambda(i|\zeta_n)$ we indicate the (conditional) logit kernel and $n(\zeta, I_M)$ is the joint density of ζ, given by the product of M standard normal densities. The unconditional probability is approximated by the following unbiased, smooth and tractable simulator:

$$\hat{P}(i) = \frac{1}{D} \sum_{d=1}^{D} \Lambda(i|\zeta_n^d) \tag{3.13}$$

where $\Lambda(i|\zeta_n^d)$ denotes the d^{th} draw from the distribution of ζ.

3.7 Taste variations

The behavior of individuals varies across the population. The simplest case arises when segments in the population are deterministically identified by the vector of the socio-economic characteristics. The definition of the segment associated to a socio-economic variable is direct when they have discrete values (for example, a *race* variable), while it requires to define classes in the continuous case. An example referring to pedestrians might be the speed of individuals. Arbitrary limiting values should be defined in order to identify, for example, slow, normal and fast pedestrians. Moreover, using this technique, different parameters might be associated to the corresponding segments, increasing the complexity of the estimation process. These cases of deterministic taste variations can be captured with a MNL model, with the standard linear-in-parameters formulation.
The fact that a parameter β varies with a continuous socio-economic variable x can be modeled as follows:

$$\beta = \hat{\beta} \left(\frac{x}{x_{ref}} \right)^{\lambda} \tag{3.14}$$

where x_{ref} is an arbitrary reference value, $\hat{\beta}$ is the parameter value associated with $x = x_{ref}$ and λ is the elasticity of the parameter with respect to x. We use this non-linear specification in Chapter 4.2

and Chapter 4.3, modeling the free-flow acceleration behavior. More complex is the case where it is not possible to segment the population in a deterministic way. The proposed solution is a *random coefficients* specification, which assumes the vector of the parameters β as randomly distributed over the population. It is possible to show that such a formulation is mathematically equivalent to the error component specification. We do not deal with this model in this work, and we remind the interested reader to McFadden and Train, 2000; Train, 2003; Hess et al., 2005.

3.8 Heteroscedasticity

The MNL model (and the GEV models in general) does not support the heteroscedasticity. Let us assume to have two different sources of data, and assume that the utilities for the alternative i, as perceived by decision makers in the first dataset, is

$$U_{in} = V_{in} + \varepsilon_{in} \tag{3.15}$$

and for those of the second dataset is

$$U_{im} = V_{im} + \varepsilon_{im} \tag{3.16}$$

where $Var(\varepsilon_{in}) \neq Var(\varepsilon_{in})$, i.e. the model is heteroscedastic. It corresponds to have an $\alpha > 0$ and $\alpha \neq 1$ such that

$$\begin{aligned} U_{in} &= V_{in} + \varepsilon_{in} \\ \alpha U_{im} &= \alpha V_{im} + \alpha \varepsilon_{im} \end{aligned} \tag{3.17}$$

with $Var(\varepsilon_{in}) = \alpha^2 Var(\varepsilon_{im})$. Now the resulting model is homoscedastic. We can say that the scale parameter α (which is unknown and has to be estimated) defined for one of the two groups, captures the heteroscedasticity in the population. This fact holds for any number G of groups, defining $G-1$ scale parameters. We use this method in Chapter 4.3, estimating the scale parameter for the Swiss portion of the dataset, with respect to the Japanese one.

3.9 Mixed GEV models

Mixed GEV models represent an extension of the model reported in Section 3.6. The idea to separate the error term into normal and Gumbel terms is still valid. The difference here is that instead of assuming i.i.d. Gumbel terms, reproducing a MNL kernel, some assumptions on the correlation patterns are made, reproducing a more general and flexible GEV kernel. The normal distributed error components remain, in order to capture additional inter-alternative correlations. Actually, all the correlation structures could be designed using the error component formulation (MNL kernel). The problem is the fact that we might have an excessive number of random terms, making the model not operational. The use of mixed GEV models is justified by the need to reduce the number of error terms, leaving to the GEV kernel the task to capture some of the correlation patterns. We use a mixed GEV model (mixed nested logit model) in Chapter 4.2.
Another possibility consists in the combination of a GEV kernel with a random coefficient specification. The aim here is to try to capture the inter-alternative correlation through the GEV kernel and the random taste variations through the random coefficients specification. The interested reader can find more details on these topics in Bath and Guo (2004), Chernew et al. (2003) and Hess et al.

(2005).

We conclude this chapter mentioning that the freeware package BIOGEME (Bierlaire, 2003) allows for the maximum likelihood estimation of the models presented here, with both linear and non-linear in parameters specifications for the utility functions.

Chapter 4

Pedestrian Walking Behavior

In Chapter 2 a review of the pedestrian modeling literature has been done. Some similar problems to those of traffic modeling are arisen, representing actually the motivation and the source of inspiration for our approach based on discrete choice analysis. In Chapter 3 we have described the theory behind DCM.

In this chapter we enter into the details of the model specification. The destination and the route are assumed to be known, generated by one of the many models listed in Chapter 1. We are interested in modeling the short range walking behavior, as a response to her immediate environment and to the presence of other pedestrians (Antonini et al., 2004a).

4.1 Adaptive spatial discretization

The interpretation of human trajectories as a sequence of choices imposes to define the first modeling element: a suitable spatial discretization. One recognized limit in most of the previous pedestrian models is that they assume a static spatial discretization. Human walking behavior is a very complex task and of course some simplifications are necessary. However, having an adaptive space model, different for each individual in the scene, allows the relaxation of (at least) part of such simplistic assumptions. Different individuals have different perception of the space around them. This intuitive idea is confirmed by numerous studies in architecture (Penn and Turner, 2002; Hillier and Hanson, 1984) and geography (Golledge, 1993; Golledge, 1999).

At a given point in time, we model where the pedestrian decide to be in a time horizon t. Typically, t is of the order of 1 second. The representation of the physical space plays an important role in the definition of the behavioral model. In our approach, we use a *dynamic and individual-based* spatial discretization representing the physical space. The basic elements that we use to define our spatial structure are illustrated in figure Figure 4.1.
The current position of the decision maker n is p_n, her current speed $v_n \in I\!R$, her current direction is $d_n \in I\!R^2$ (normalized, so that $\|d_n\| = 1$) and her visual angle is θ_n. The region of interest is situated in front of the pedestrian, within her visual field represented by the shaded area in Figure 4.1. The size of the region of interest is conceptually given by the depth of decision maker visual field. In practice, being the dimension of the analyzed scene reasonable, we have simply considered the

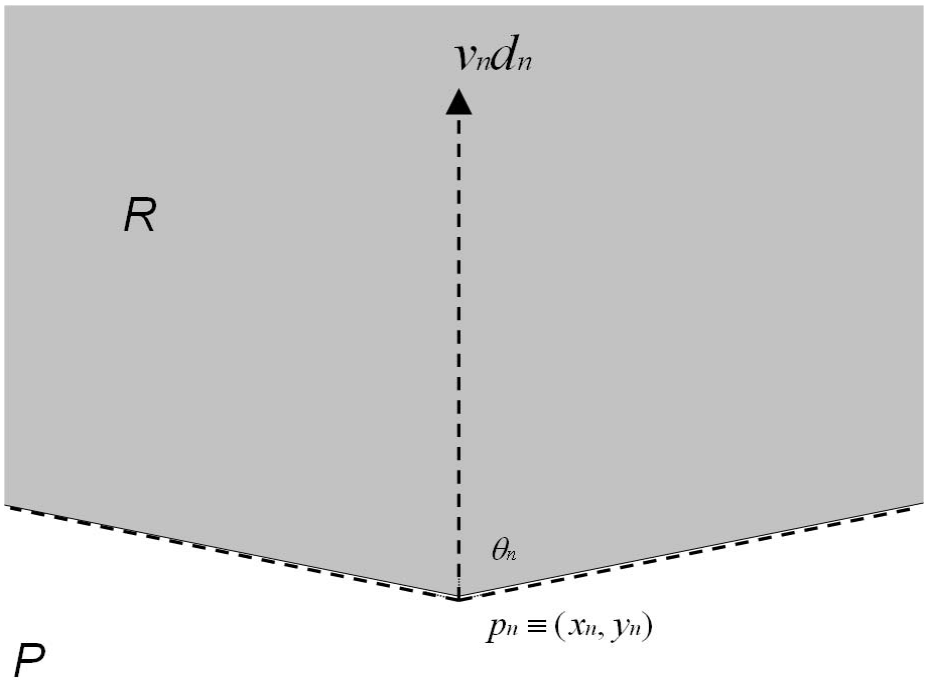


Figure 4.1: The basic geometrical elements of the space structure

region of interest coinciding with the scene floor size.

The choice set consists of a combination of speed regimes and directions. With regard to speed regimes, the decision-maker has three possibilities: keep the same speed v_n, slow down to $v_{dec} = (1-\gamma)v_n$ or accelerate up to $v_{acc} = (1+\gamma)v_n$, where v_n is the current speed of the decision maker and γ an accelaration/deceleration factor. In our model, we have arbitrarily selected $\gamma = 0.5$. With regard to directions, the visual angle $\theta_n = 170°$ (Costella, 1992) is segmented into 11 radial cones, one cone capturing the decision not to change the direction (assumed to have an angle of 10°), and 10 cones capturing the decision to change direction, 5 at the left of the central cone, and 5 symmetrically defined at the right, as illustrated in Figure 4.2. Note that the apertures of those cones are not equal. Cones far from the central one have larger angles, as mentioned in Figure 4.2. Each cone is characterized in the model by its bisecting direction, denoted by d and assumed to be normalized, that is $\|d\| = 1$. The central cone is obviously characterized by the current direction d_n. Each alternative with speed v and direction d is characterized by the physical center of the corresponding cell in the space discretization c_{vd}, that is

$$c_{vd} = p_{\mathrm{n}} + vtd.$$

The above assumptions arise from a twofold justification. First, we need to calibrate models where the number of alternatives plays an important role in the computational efficiency of the estimation process. In this spirit, 3 speed regimes and 11 radial directions represent a good tradeoff between choice set size and flexibility in the captured behaviors. Second, we are interested in direction changes and speed variations determined by actual choices. By using a higher spatial resolution the risk is to have numerous choices not corresponding to actual behaviors. It is important to emphasize that this conceptual universal choice set, composed of $N = 33$ alternatives, is associated with different

physical locations in space, depending on the current position and speed of the decision-maker. We refer to it as a *dynamic* and *individual-based* spatial discretization. A universal choice set derived from a simple static discretization of the space, similar to the model used by CA approaches, would have been too cumbersome and not sufficiently flexible in a discrete choice context.

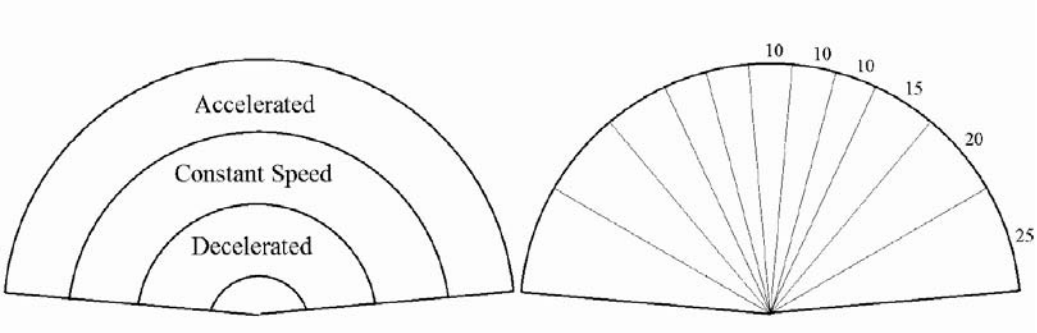


Figure 4.2: Choice set

For each individual, some cells can be declared unavailable because there is a physical obstacle blocking the corresponding space. Also, a maximum speed can be assigned to each individual (it can be fixed for the entire population, or drawn from a distribution). If the pedestrian is already walking at maximum speed, the cells corresponding to acceleration are declared not available.

4.2 A first model specification

We denote by c_{vdn} the alternative of individual n corresponding to speed regime $v \in \{v_{\rm n}, v_{\rm dec}, v_{\rm acc}\}$, and direction d. The utility associated with this alternative is a random variable, for which the deterministic part is defined as

$$\begin{array}{rcll} V_{vdn} & = & \beta_{\rm occ} & {\rm occupation}_{vd} & + \\ & & \beta_{\rm dir} & {\rm dir}_{dn} & + \\ & & \beta_{\rm ddir} & {\rm ddir}_{dn} & + \\ & & \beta_{\rm angle} & {\rm angle}_{vdn} & + \\ & & \beta_{\rm acc} & I_{\rm v,acc}(v_n/v_{\rm max})^{\lambda_{acc}} & + \\ & & \beta_{\rm dec} & I_{\rm v,dec}(v_n/v_{\rm max})^{\lambda_{dec}} & \end{array} \tag{4.1}$$

where $\beta_{\rm occ}$, $\beta_{\rm dir}$, $\beta_{\rm dest}$, $\beta_{\rm angle}$, $\beta_{\rm acc}$, λ_{acc}, $\beta_{\rm dec}$, and λ_{dec} are unknown parameters to be estimated. Note that this specification is the result of an intensive modeling process, where many different specifications have been tested. The attributes describe the environment of the decision-maker. Namely, the position and direction of other pedestrians are important. We assume that there are N pedestrians potentially influencing the decision-maker, ideally all those staying inside the visual field of the decision maker. Each pedestrian k is at a position p_k and walks toward a direction d_k. The attributes are defined as follows:

occupation$_{vd}$ It is defined as the weighted number of pedestrians being in the cone characterized by d, that is

$$ {\rm occupation}_{vd} = \sum_{k=1}^{N} I_{kd} e^{-\gamma_1 \|p_k - c_{vdn}\|} \tag{4.2}$$

where N is the total number of pedestrians in the environment, I_{kd} is one if pedestrian k belongs to the cone characterized by d and 0 otherwise, $\|p_k - c_{vdn}\|$ is the distance between pedestrian k and the physical center of the alternative c_{vdn}. The role of γ_1 is to weight the importance of the distance in the formula. It is designed to capture the influence of the proximity of other pedestrians in the movements decisions as illustrated in Figure 4.3. We arbitrarily fix the value of γ_1 equal to 1.

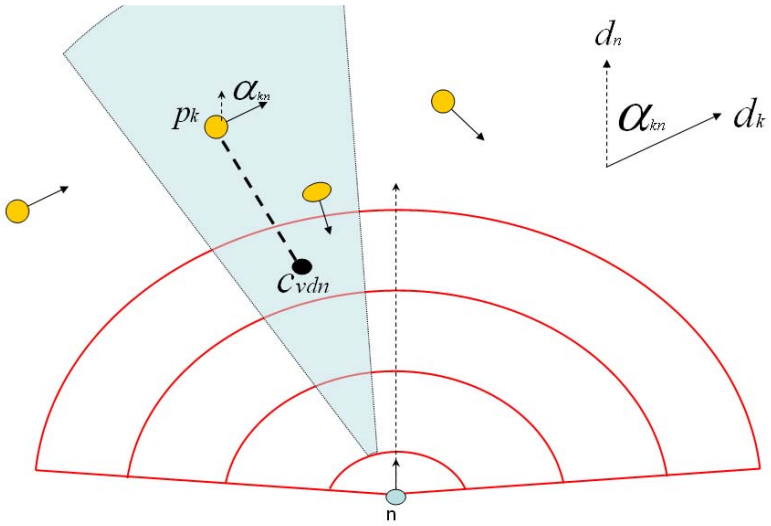


Figure 4.3: Occupation and angle

dir$_{dn}$ It is defined as the angle (in degrees) between direction d and direction d_n, corresponding to the central cone, as shown in Figure 4.8. It is designed to capture the propensity of pedestrians to prefer their current direction, and not to erratically modify it.

ddir$_{dn}$ If we denote by D_n the direction pointing toward the actual destination of decision-maker n, this attribute is defined as the angle in degrees between D_n and d, as shown in Figure 4.8. It is designed to capture the propensity of pedestrians to move toward their destination.

angle$_{vdn}$ It is defined as the weighted sum of the angles between direction d_n and the walking directions of other pedestrians, that is

$$\text{angle}_{vdn} = \sum_{k=1}^{N} I_{kd}\alpha_{kn}e^{-\gamma_2\|p_k - c_{vdn}\|} \tag{4.3}$$

where α_{kn} is the angle between d_n and d_k (Figure 4.3). The indicator function I_{kd} equals to 1 if pedestrian k is inside the cone characterised by direction d. The role of γ_2 is similar to the role of γ_1 in Eq. (4.2). This attribute is designed to capture the influence of the other pedestrians dynamics. Indeed, if a pedestrian k walks in the same direction (angle=0) or in the opposite direction (angle=π) with respect to the decision maker, we expect a different influence on the choice of the decision-maker. Note that, similarly to the definition of occupation$_{vd}$, close pedestrians play a more important influence than those who are further away. In our tests, we have arbitrarily fixed γ_2 to 1.

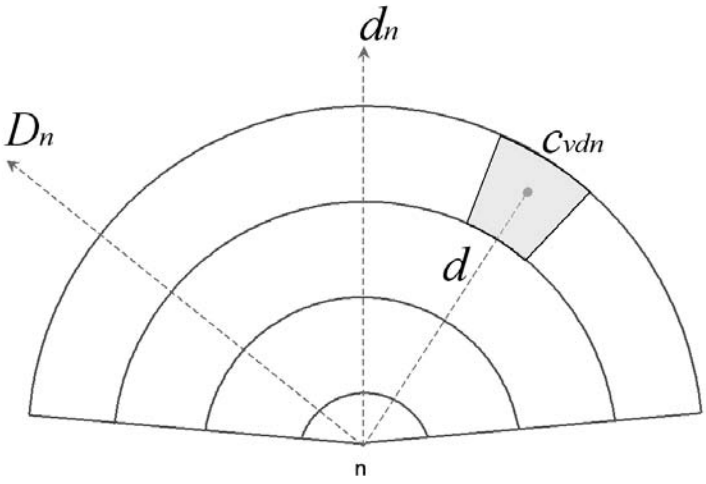


Figure 4.4: Destination and direction

We finally comment on the last two terms of the utility function Eq. (4.1). The attribute $I_{\mathrm{v,acc}}$ is 1 if $v = v_{\mathrm{acc}}$, that is, if the alternative corresponds to an acceleration and 0 otherwise. $I_{\mathrm{v,dec}}$ is similarly defined. If we write

$$\begin{aligned} \tilde{\beta}_{\mathrm{acc}} &= \beta_{\mathrm{acc}}(v_n/v_{\max})^{\lambda_{acc}}, \\ \tilde{\beta}_{\mathrm{dec}} &= \beta_{\mathrm{dec}}(v_n/v_{\max})^{\lambda_{dec}}, \end{aligned}$$

the parameters $\tilde{\beta}_{\mathrm{acc}}$ and $\tilde{\beta}_{\mathrm{dec}}$ are simple dummies for the acceleration and deceleration alternatives, respectively, capturing the attractiveness of acceleration, respectively deceleration. We postulate that these dummies vary with the current speed of the decision-maker v_n. Indeed, someone who is already walking fast has less incentive to accelerate than someone who is walking slowly. The value of the parameter $v_{\max}$ is arbitrary. We have set it to the maximum speed observed in the data, for numerical convenience. β_{acc} is the value of the dummy associated with $v_n = v_{\max}$ and λ_{acc} is the elasticity of the dummy with respect to speed, that is

$$\lambda_{acc} = \frac{\partial \tilde{\beta}_{\mathrm{acc}}}{\partial v_n} \frac{v_n}{\tilde{\beta}_{\mathrm{acc}}} \tag{4.4}$$

We conclude this section pointing out the fact that, with the exception of the current speed value, the variables introduced in the utility functions are decision maker independent. This fact represents of course a limitation in the explanatory power of the model but it is dictated by reasons related to the available data. They are actually collected from video sequences, so no socio-economic characteristics are available. A detailed description of the collected data is given in Chapter 5.1.

4.2.1 The random variable

As introduced in Chapter 3, the utility of each alternative is a random variable containing a systematic part and a random part. Different assumptions about the random term give rise to different models. We present in this Section two different model formulations: a cross nested logit model and a nested logit with error components.

Cross nested logit formulation

The theoretical aspects of the CNL model have been described in Section 3.5.1. We assume a correlation structure depending on the speed and direction and we identify five nests in the choice set: *accelerated*, *constant*, *decelerated*, *central* and *not central*. This correlation structure is illustrated in Figure 4.5. The degrees of membership to the different nests (α_{jm}, see Eq. (3.9)) are fixed to the constant value 0.5. This choice reflects the lack of a precise prior knowledge on such coefficients. In principle, it would be possible also to estimate them from the data, increasing considerably the complexity of the model as well as the computational time of the estimation. For identification purposes, the global scale of the model μ is fixed to 1 and we estimate the scale parameters for the error terms associated with the five nests, μ_m.

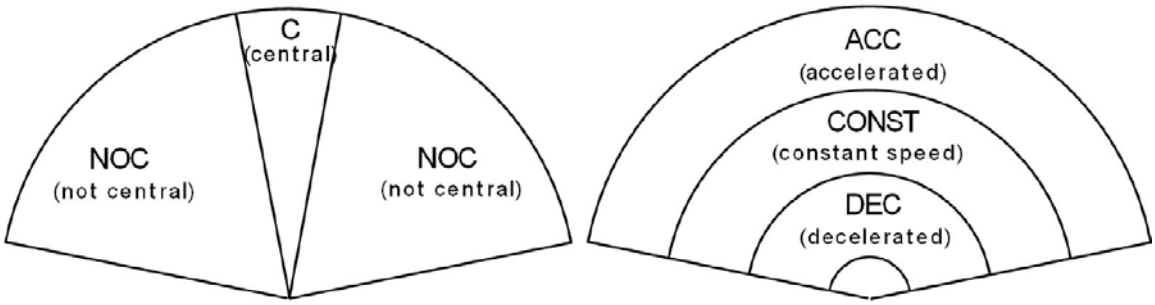


Figure 4.5: left: Nesting based on direction **right:** Nesting based on speed

Nested logit with error components

The theoretical aspects of the Logit Kernel and mixed GEV models have been illustrated in Section 3.6 and Section 3.9, respectively. Here we specify a nested logit model with an error components formulation, where the correlation between alternatives still depends on speed and direction. Three Extreme Value error terms capture the correlation in the speed related nests (accelerated, constant and decelerated), while 11 normally distributed error components capture the correlation between alternatives along the 11 radial directions, one component for each direction. We show this structure in Figure 4.6. The mixed NL formulation only changes the error structure of the model, with respect to the previous CNL formulation. The systematic part of the utility functions is exactly the same as before.

The utility function for alternative i as perceived by the individual n has the following form:

$$U_{in} = V_{in} + \xi_k + \varepsilon_{in} \tag{4.5}$$

where V_{in} is defined by (Eq. (4.1)), ξ_k are normally distributed error components with zero mean and variance σ_k, $k = 1, ..., 11$ the index of directions, ε_{in} is the error term capturing the nested structure, $n = 1, ..., N$ is the index of alternatives. If the ξ_k are known, the model corresponds to a NL formulation:

$$P(i|\xi_k) = \frac{e^{\mu_{s_{(i)}} V_i}}{\sum_{j \in s_{(i)}} e^{\mu_{s_{(i)}} V_j}} \cdot \frac{e^{\mu \bar{V}_{s_{(i)}}}}{\sum_{s \in S} e^{\mu \bar{V}_s}} \tag{4.6}$$

where

$$\bar{V}_s = \frac{1}{\mu_s} \ln \sum_{j \in s} \exp \mu_s V_j \tag{4.7}$$

and $P(i|\xi_k)$ is the probability that the choice is i conditional to ξ_k. The index $s_{(i)}$ refers to the nest s containing the alternative i, where $s \in S = \{accelerated, constant, decelerated\}$. The μ and μ_s are the global and the nest scale factors, respectively. Since the ξ_k are unknown, the unconditional choice probability is given by

$$P(i) = \int_{\xi} P(i|\xi_k) n(\xi, I_M) d\xi \tag{4.8}$$

with $n(\xi, I_M)$ being the joint density of ξ given by the product of $M = 11$ standard normal densities. The unconditional probability is approximated by the following unbiased, smooth and tractable simulator:

$$\hat{P}(i) = \frac{1}{D} \sum_{d=1}^{D} P(i|\xi_k^d) \tag{4.9}$$

where $P(i|\xi_k^d)$ denotes the d^{th} draw from the distribution of ξ.

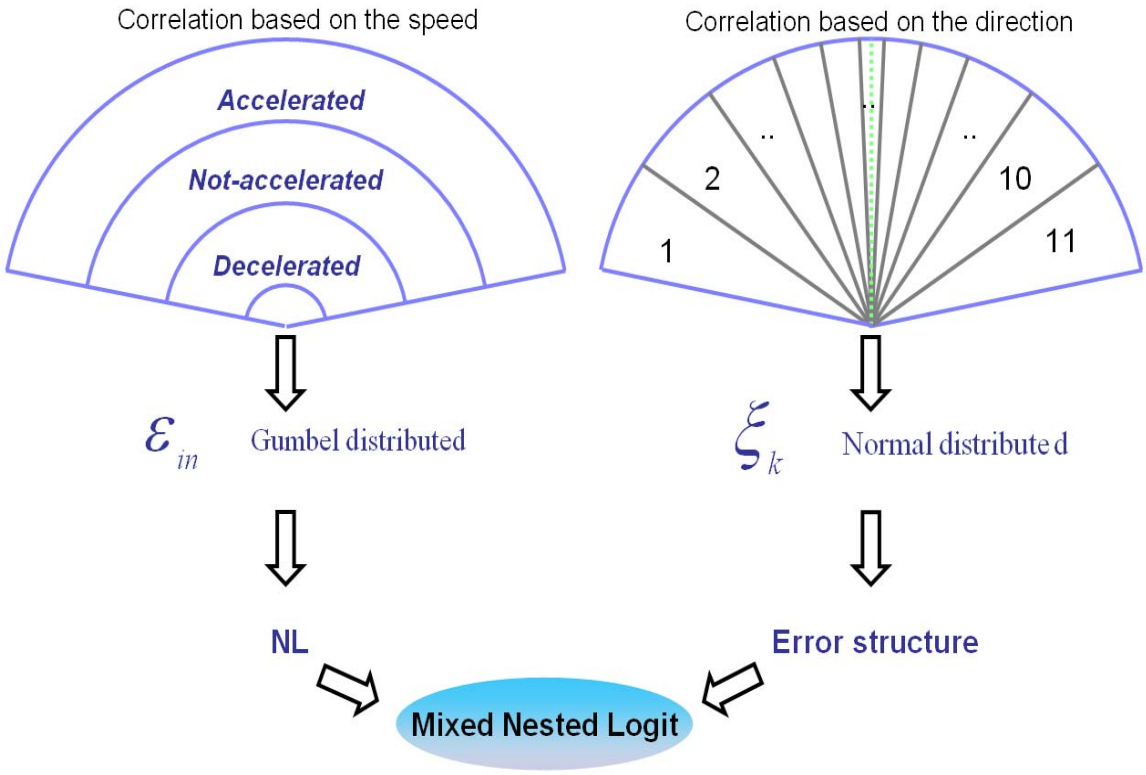


Figure 4.6: Correlation structure in the Mixed Nested Logit formulation

Assuming such an error structure, the parameters that have to be estimated are the scale parameters μ_s of the extreme value error terms associated with the speed-based nests and the 11 standard deviations σ_k of the normal error components associated with the directions.

4.3 A more general framework

In this Section more insights are gained on the interactions between individuals. We keep here the same space model (leading to the same choice set definition) and the cross nested logit error structure presented in Section 4.2. A more general framework for pedestrian walking behavior is introduced, while detailed models for pedestrian-pedestrian interactions are specified (Antonini and Bierlaire, 2005).

4.3.1 Behavioral framework

We refer to the general framework for pedestrian behavior described by Hoogendoorn (2003) and Daamen (2004). Individuals in a certain environment make different decisions, following a hierarchical scheme: *strategical*, *tactical* and *operational*. We are interested in modeling the operational level of such a hierarchy. Two types of behavior are modeled here: *unconstrained* and *constrained*. The constrained patterns are further classified into *attractive interactions* and *repulsive interactions*. This conceptual framework is illustrated in Figure 4.7.

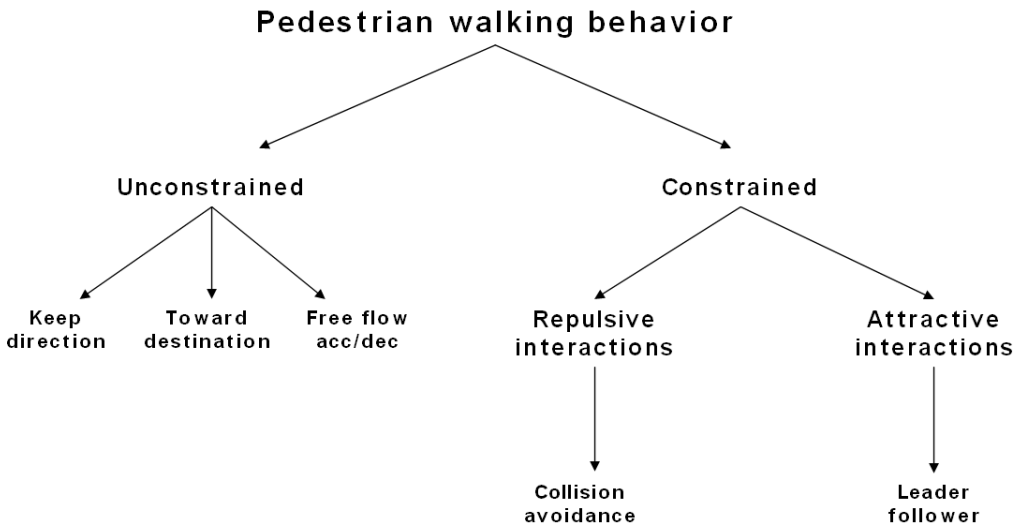


Figure 4.7: Conceptual framework for pedestrian walking behavior

The unconstrained decisions are independent from the presence of other pedestrians and are generated by subjective and/or unobserved factors. The first of these factors is represented by the individual's destination. It is assumed to be exogenous to the model and decided at the strategical level. The second factor is represented by the tendency of people to keep their current direction, minimizing their angular displacement. Finally, unconstrained accelerations (with accelerations we mean both positive and negative speed variations) are dictated by the individual desired speed. The implementation of these ideas is made through the three unconstrained patterns indicated in Figure 4.7.

The main contribution of this chapter consists in a detailed analysis of the constrained behaviors. We assume that behavioral constraints are induced by the interactions with the other individuals in the scene. Repulsive interactions are modeled through the *collision avoidance* pattern, which is

designed to capture the effects of possible collisions on the current trajectory of the decision maker. Attractive interactions are modeled through the *leader-follower* behavior, that is the tendency of people to follow another individual in a crowd, in order to benefit from the space she is creating. Here, the existence of one or more leaders is assumed. They are represented by those individuals in a neighbor of the decision maker and with similar moving directions and speed, affecting her decisions.

4.3.2 Assumptions

Individuals walk on a 2D plane and any kind of behavior influencing their movement results in two kind of observations: changes in direction and changes in speed, i.e. accelerations. This specification is important to perform walking behavior analysis, and hypotheses have to be made concerning the unobserved factors in the model and how they are related to the observed data. Figure 4.7 summarizes the set of assumptions we want to test. Five behavioral patterns are defined. In a discrete choice context, they have to be considered as competitive terms entering the utility functions of each alternative, as reported in Equation 4.10. The utilities describe the space around the decision maker and under the rational behavior assumption the individual will choose that location (alternative) with the maximum utility. In the following, we discuss the different patterns and the associated assumptions in more details.

Unconstrained patterns

The unconstrained patterns are identified by those behaviors that are independent from the presence of other pedestrians. In this conditions we assume that three factors influence the individual behavior.

- **Toward destination** The first factor is represented by the choice of the final destination which can be a specific area where the individual wants to perform a certain activity in her schedule. We assume that the destination choice is performed at the strategical layer in the hierarchical decision process. Such a higher level choice is naturally reflected on the short term behavior as the tendency of individuals to choose, for the next step, such spatial location that minimize both the angular displacement and the distance from the destination.
- **Keep direction** The second factor influencing the unconstrained behavior is represented by the tendency of people to avoid frequent changings in direction. People choose their next position in order to minimize the angular displacement from their current movement direction, analogously to the specification presented in Section 4.2. In addition to the behavioral motivation of this factor, it also plays a smoothing role in the model, avoiding drastic changes of direction from one time period to the next.
- **Free flow acceleration** In free flow conditions the behavior of the individual is driven by her desired speed. The acceleration is then a function of the difference between current speed and desired speed. However, this factor is an unobserved individual characteristic and it can not be introduced explicitly in the model. The specification for this pattern is the same as that presented in Section **??**spec1).

Constrained patterns

Constrained behaviors are induced by the presence of other individuals in the scene and capture the pedestrian-pedestrian interactions. We identify attractive and repulsive interactions, described by the following patterns.

- **Leader-follower** We assume that the decision maker is influenced by leaders. In our spatial representation 11 radial cones partition the choice set (see Figure 4.2). In each of these directions a possible leader can be identified among a set of *potential leaders*. A potential leader is an individual which is inside a certain region of interest, *not so far* from the decision maker and having a moving direction *close enough* to the direction of the radial cone where she is. Once identified, the leader induces an attractive interaction on the decision maker. Similarly to car following models, a leader acceleration (deceleration) corresponds to a decision maker acceleration (deceleration).

- **Collision avoidance** This pattern captures the effects of possible collisions on the decision maker trajectory. For each direction in the choice set, a collider is identified among a set of *potential colliders*. Another individual is selected as a potential collider if she stays inside a certain region of interest, *not so far* from the decision maker and walking against the decision maker herself. This pattern is associated with repulsive interactions in the obvious sense that pedestrians change their current direction avoiding collisions with other individuals.

4.3.3 The model

Following the framework proposed in Figure 4.7 we report here the systematic utilities of the model:

$$\left.\begin{aligned} V_{vdn} = \beta_{\text{dir}}\text{dir}_{dn} \quad + \\ \end{aligned}\right\} \quad \textit{keep direction}$$

$$\left.\begin{aligned} \beta_{\text{ddist}}\text{ddist}_{vdn} \quad + \\ \beta_{\text{ddir}}\text{ddir}_{dn} \quad + \\ \end{aligned}\right\} \quad \textit{toward destination}$$

$$\left.\begin{aligned} \beta_{\text{acc}} I_{\text{v,acc}} (v_n / v_{\max})^{\lambda_{acc}} \quad + \\ \beta_{\text{dec}} I_{\text{v,dec}} (v_n / v_{\max})^{\lambda_{dec}} \quad + \\ \end{aligned}\right\} \quad \textit{free flow acceleration} \tag{4.10}$$

$$\left.\begin{aligned} I_{\text{v,acc}} I^L_{acc} \alpha^L_{acc} D_L^{\rho^L_{acc}} \Delta v_L^{\gamma^L_{acc}} \Delta\theta_L^{\delta^L_{acc}} \quad + \\ I_{\text{v,dec}} I^L_{dec} \alpha^L_{dec} D_L^{\rho^L_{dec}} \Delta v_L^{\gamma^L_{dec}} \Delta\theta_L^{\delta^L_{dec}} \quad + \\ \end{aligned}\right\} \quad \textit{leader follower}$$

$$\left.\begin{aligned} I_{\text{d},d_n} I_C \alpha_C e^{-\rho_C D_C} \Delta v_C^{\gamma_C} \Delta\theta_C^{\delta_C} \quad \\ \end{aligned}\right\} \quad \textit{collision avoidance}$$

where all the β parameters as well as λ_{acc}, λ_{dec}, α^L_{acc}, ρ^L_{acc}, γ^L_{acc}, δ^L_{acc}, α^L_{dec}, ρ^L_{dec}, γ^L_{dec}, δ^L_{dec}, α_C, ρ_C, γ_C, δ_C are unknown and have to be estimated. Note that this specification is the result of an intensive modeling process, where many different specifications have been tested. We explain in the following the different terms of the utilities.

- **Keep direction** This behavior is captured by the term $\beta_{dir} dir_{dn}$ as in the first model specification (see Equation 4.1 and Figure 4.8).

- **Toward destination** This behavior is slightly modified with respect to the first specification. It is captured by the term

$$\beta_{ddist} ddist_{vdn} + \beta_{ddir} ddir_{dn}$$

where the variable $ddir_{dn}$ is the same as in Equation 4.1 and $ddist_{vdn}$ is defined as the distance between the destination and the center of the alternative C_{vdn}. The direction and distance attributes for the destination are introduced in order to take into account the two dimensional nature of pedestrian mevements (see figure 4.8).

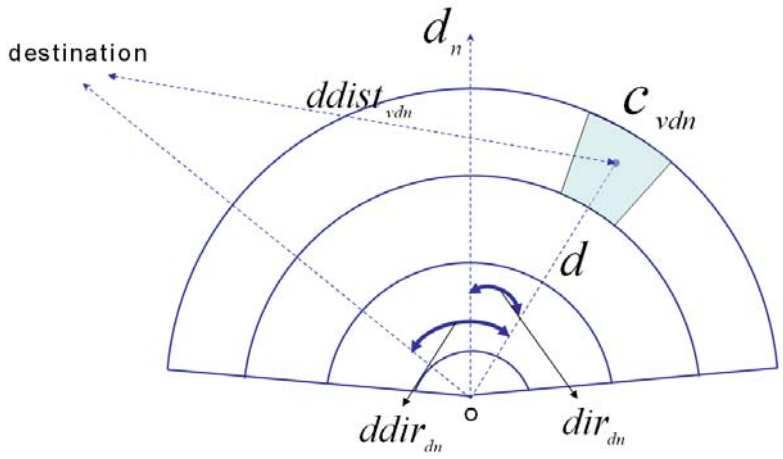


Figure 4.8: The elements capturing the *keep direction* and *toward destination* behaviors

- **Free flow acceleration** We keep here the same model used in Equation 4.1 for the acceleration and deceleration dummies.
- **Leader-follower** The leader-follower model captures the attractive interactions among pedestrians and is given by the following terms

$$I_{\mathrm{v,acc}} I^L_{acc} \alpha^L_{acc} D_L^{\rho^L_{acc}} \Delta v_L^{\gamma^L_{acc}} \Delta\theta_L^{\delta^L_{acc}} + I_{\mathrm{v,dec}} I^L_{dec} \alpha^L_{dec} D_L^{\rho^L_{dec}} \Delta v_L^{\gamma^L_{dec}} \Delta\theta_L^{\delta^L_{dec}}.$$

It is described by a *sensitivity/stimulus* framework. For a given leader, the sensitivity is described by

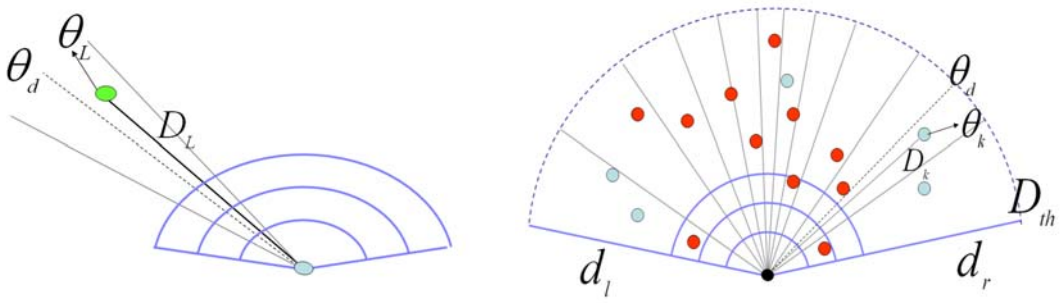
$$sensitivity = f(D_L) = \alpha^L_g D_L^{\rho^L_g} \tag{4.11}$$

where D_L represents the distance between the decision maker and the leader. The parameters α^L_g and ρ^L_g have to be estimated and $g = \{acc, dec\}$ indicates when the leader is accelerating or decelerating with respect to the decision maker. The decision maker reacts to stimuli coming from the chosen leader. We model the stimulus as a function of the leader's relative speed Δv_L and the leader's relative direction $\Delta\theta_L$ as follows:

$$stimulus = g(\Delta v_L, \Delta\theta_L) = \Delta v_L^{\gamma^L_g} \Delta\theta_L^{\delta^L_g} \tag{4.12}$$

with $\Delta v_L = |v_L - v_n|$, where v_L and v_n are the leader's speed module and the decision maker's speed module, respectively. The variable $\Delta\theta_L = \theta_L - \theta_d$, where θ_L represents the leader's movement direction and θ_d is the angle characterizing direction d, as shown in Figure 4.9(a). The parameters γ^L_g and δ^L_g have to be estimated. A leader acceleration (deceleration) induces a decision maker's acceleration (deceleration). A substantially different movement direction in the leader reduces the influence of the latter on the decision maker.

The leader for each direction is chosen considering several *potential leaders*, as shown in Figure 4.9(b). An individual k is defined as a potential leader based on the following indicator function I^k_g:



(a) (b)

Figure 4.9: Figure 4.9(a) shows the leader's movement direction, θ_L, the direction of the radial cone where the leader lies, θ_d, and her distance from the decision maker, D_L, used in the definitions of both the sensitivity and the stimulus terms. Figure 4.9(b) illustrates how many potential leaders are considered for each direction and how only the nearest one is chosen as leader for a specific direction (darker circles)

$$I_g^k = \begin{cases} 1, & \text{if } d_l \leq d_k \leq d_r, \\ & \text{and } 0 < D_k \leq D_{th}, \\ & \text{and } 0 < |\Delta\theta_k| \leq \Delta\theta_{th} \\ 0, & \text{otherwise.} \end{cases}$$

where d_l and d_r represent the bounding left and right directions of the choice set (defining the region of interest) while d_k is the direction identifying the pedestrian k position. D_k is the distance between pedestrian k and the decision maker, $\Delta\theta_k = \theta_k - \theta_d$ is the difference between the movement direction of pedestrian k (θ_k) and the angle characterizing direction d, i.e. the direction identifying the radial cone where individual k lies (θ_d). The two thresholds D_{th} and $\Delta\theta_{th}$ are fixed at the values $D_{th} = 5D_{max}$, where D_{max} is the radius of the choice set, and $\Delta\theta_{th} = 10$ degrees. We assume an implicit *leader choice* process, executed by the decision maker herself and modeled choosing as leader for each direction the potential leader at the minimum distance $D_L = \min_{k\in K}(D_k)$, illustrated in Figure 4.9(b) by the darker circles. Finally, the indicator functions $I_{v,acc}$ and $I_{v,dec}$ discriminate between accelerated and decelerated alternatives, as for the free flow acceleration model.

- **Collision avoidance** The collision avoidance model captures the repulsive interactions among pedestrians and is given by the following term

$$I_{\mathrm{d},d_n} I_C \alpha_C e^{-\rho_C D_C} \Delta v_C^{\gamma_C} \Delta\theta_C^{\delta_C}$$

The scenario is similar to the leader follower. We keep the sensitivity/stimulus framework, where the sensitivity function is defined as

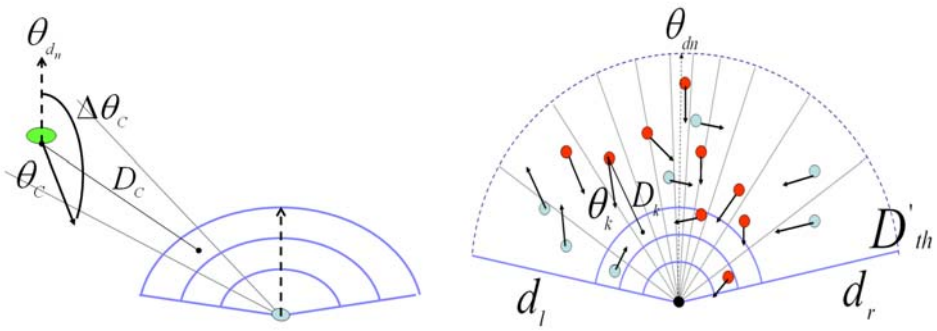
$$sensitivity = f(D_C) = \alpha_C e^{-\rho_C D_C} \tag{4.13}$$

where the parameters α_C and ρ_C have to be estimated and D_C is the distance between the collider position and the center of the alternative, as shown in Figure 4.10(a). We choose the

exponential to keep the same functional form as that used in Antonini et al. (2004b). The decision maker reacts to stimuli coming from the collider. We model the stimulus as a function of two variables:

$$stimulus = f(\Delta v_C, \Delta\theta_C) = \Delta v_C^{\gamma_C} \Delta\theta_C^{\delta_C} \tag{4.14}$$

with $\Delta\theta_C = \theta_C - \theta_{d_n}$, where θ_C is the collider movement direction and θ_{d_n} is the decision maker movement direction, and $\Delta v_C = v_C + v_n$, where v_C is the collider's speed module and v_n is the decision maker's speed module. The parameters γ_C and δ_C have to be estimated. Individuals walking against the decision maker at higher speeds and in more frontal directions (higher $\Delta\theta_C$) generate stronger reactions, weighted by the sensitivity function.



(a) (b)

Figure 4.10: Figure 4.10(a) shows the collider and decision maker movement directions, θ_C and θ_{dn} respectively. D_C represents here the distance of the collider with the center of the alternative. Figure 4.10(b) shows many potential colliders taken into account for each direction

The collider for each direction is chosen considering several *potential colliders*, as shown in Figure 4.10(b). An individual k is defined as a potential collider based on the following indicator function

$$I_C^k = \begin{cases} 1, & \text{if } d_l \leq d_k \leq d_r, \\ & \text{and } 0 < D_k \leq D'_{th}, \\ & \text{and } \frac{\pi}{2} \leq |\Delta\theta_k| \leq \pi \\ 0, & \text{otherwise.} \end{cases}$$

where d_l, d_r and d_k are the same as those defined for the leader follower model. D_k is now the distance between individual k and the center of the alternative, $\Delta\theta_k = \theta_k - \theta_{d_n}$ is the difference between the movement direction of pedestrian k (θ_k) and the movement direction of the decision maker, θ_{d_n}. The value of the distance threshold is now fixed to $D'_{th} = 10 D_{max}$. We use a larger value for such a threshold compared to the leader-follower model, assuming the collision avoidance behavior being a longer range interaction, happening also at a lower density level. We assume an implicit *collider choice* process, executed by the decision maker herself. Among the set of K_d potential colliders for direction d, the collider is chosen as that individual having $\Delta\theta_C = \max_{k \in K_d} |\Delta\theta_k|$. The related indicator function is I_C. Finally, the collision avoidance term is included in the utility functions of all the alternatives, with the

exception of the central ones. So, the indicator function I_{d,d_n} is equal to 1 for those alternatives that are not in the current direction ($d \neq d_n$), 0 otherwise.

4.4 Summary

In this chapter the pedestrian walking behavior model has been presented. An adaptive spatial discretization has been introduced, modeling the fact that different individuals differently perceive the surrounding available space, in order to move their next step. In our approach, the choice set is always oriented along the current pedestrian moving direction and its size is a function of the current individual speed module.
As already said in Chapter 1, behavioral modeling is a strong iterative process. In this spirit, two model specifications are described here, reflecting the logical order of the model development. In the first one, the focus is on the unconstrained patterns, and a first attempt is done in order to model pedestrian-pedestrian interactions. The error structure is captured by two alternative solutions, a cross nested logit and a mixed nested logit models. The second specification provides a more general framework, where the unconstrained patterns are slightly changed, focussing on the development of the constrained behaviors, resulting in the leader follower and collision avoidance models. This model is more sophisticated than the previous one, presenting non-linearities in most of the included patterns. The error structure here is captured with the cross nested logit specification, also used in the first model. The estimation results for both the specifications are reported and discussed in Section 5.2.

Chapter 5

Estimation results and validation

In this chapter we describe the data used for model calibration in Section 5.1 and we report the estimation results for the different specifications and error structures described in the previous chapter in Section 5.2. Finally, in order to validate the model, a pedestrian simulator is described and a simulation case is presented.

5.1 Collected data

We have used two datasets for the estimation of the unknown parameters. They have been collected manually tracking pedestrians in video sequences. In this chapter we describe the collection method and we provide descriptive statistics for both the collected datasets.

5.1.1 The Swiss dataset

This dataset has been collected from digital video sequences of actual pedestrians recorded in the city of Lausanne in 2002, next to an entrance of a metro station. An image of the corresponding scenario is shown in Figure 5.1.

The fundamental condition for our data collection process is the calibration of the video camera used to make the video sequences in order to match the image with the walking plane. The camera calibration is a process providing an explicit mapping (under a 3x4 projection matrix) between a real 3D world point and a 2D image pixel. From a mathematical point of view, the calibration problem gives rise to a system of non-linear equations where the variables are the following camera parameters: height, focal distance and the three angles with respect to the three camera axes. In these conditions, it is possible to associate one and only one point on the walking plane with the corresponding point on the image plane, and vice-versa. A full description of the calibration problem is out of the scope of this thesis and we refer the reader to the related literature (Heikkila, 2000, Willson, 1994 among others).

The video sequences have been converted to AVI format with a frame rate equal to 10 frames/second. We have collected data for 36 pedestrians for a total of 1675 observed positions, with a time interval of 3 frames (0.3 seconds). Having a calibrated camera, it has been possible to manually track the individuals, projecting the coordinates for each observation from the image plane to the walking plane, giving rise to trajectories.

Figure 5.1: Swiss scenario

It is important to underline the fact that the manual tracking of pedestrians introduces some systematic errors in the data collection process. First of all, it is practically difficult to track the same point (the feet in our case) on the image, especially in cases of partial and/or total occlusions between the individuals. Moreover, being the camera calibration a numerical approximation to the real solution, projection errors are present in the stored trajectories. Such errors are more accentuated for those positions at the top of the image. However, the model is based on relative movements and these errors are systematic. This allows for most errors to cancel out, with few consequences on the model estimation process.

We derive the speed data from the positions :

$$\mathbf{v}_k^{t+1} = \frac{\mathbf{p}_k^{t+1} - \mathbf{p}_k^t}{\Delta t} \tag{5.1}$$

where $\mathbf{p}_k^t$ represents the position vector for pedestrian k at time t. The motion direction information is the normalized vector:

$$\begin{pmatrix} arccos(v_{kx}^t / \|\mathbf{v}_k^t\|) \\ arccos(v_{ky}^t / \|\mathbf{v}_k^t\|) \end{pmatrix}$$

v_{kx}^t and v_{ky}^t are the observed speed vector components at time t. Table 5.1 shows some data statistics. For each pedestrian, we report the trajectory length and the average values of the speed module, the angle between two consecutive directions and the angle between the current direction and the destination (taken as the last point in the pedestrian's trajectory). The speed values are expressed in m/s and all the angles are in degrees.
We report in Figure 5.3(a), Figure 5.3(b) two examples of manually tracked trajectories. Finally, Figure 5.4(a) and Figure 5.4(b) report the speed-time graphs for the same pedestrians. Looking at the trajectory of pedestrian in (Figure 5.3(b)) we can see how this individual is moving away from the camera position. On the image plane the farthest positions are located on the top of the image where the projection errors are larger. This is reflected in the speed-time graph, where

the speed values increase over time. Figure 5.2 illustrates the speed distribution and Table 5.2 the speed statistics. We finally report in Figure 5.5 the histogram of the revealed choices. We note how such a distribution presents three modes, around the three alternatives corresponding to the central direction, as expected.

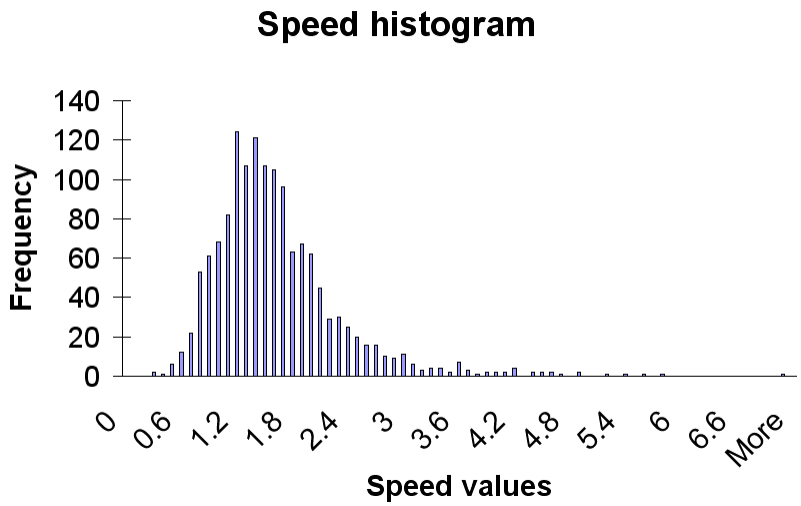
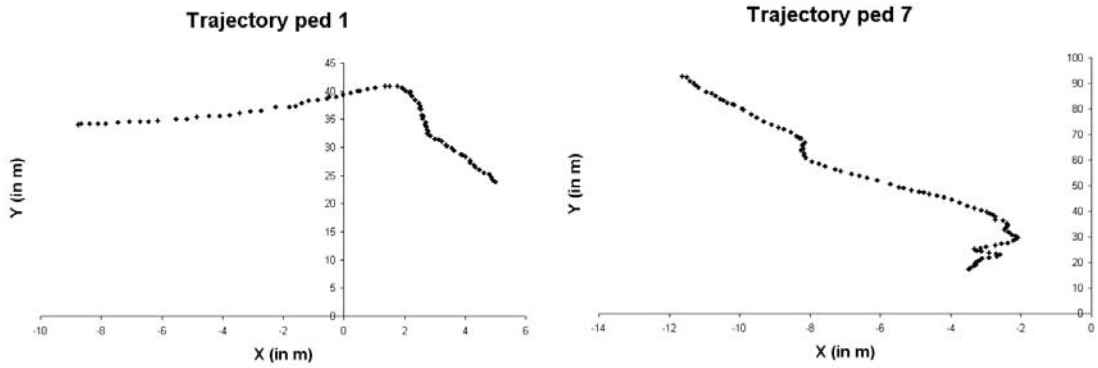


Figure 5.2: Speed histogram

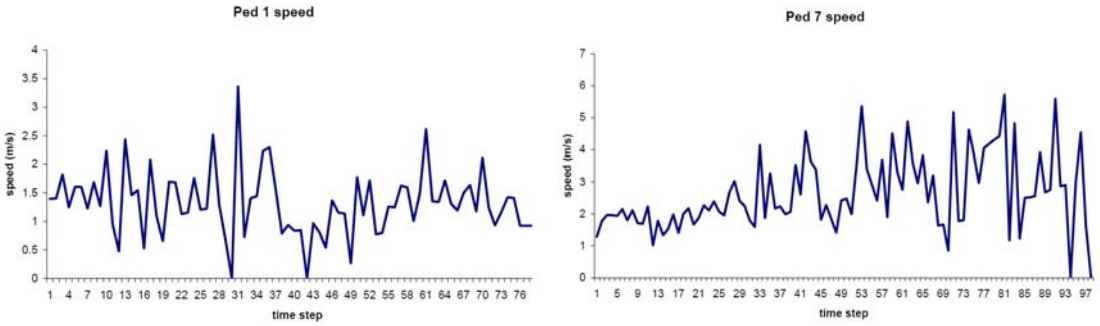


(a) (b)

Figure 5.3: Examples of two manually tracked trajectories

5.1.2 The Japanese dataset

This dataset has been collected in Sendai, Japan, on August 2000 (see Teknomo et al., 2000, Teknomo, 2002). The video sequence has been recorded from the 6th floor of the JTB parking building (around 19 meter height), situated at a large pedestrian crossing point. An image of the corresponding scenario is shown in Figure 5.6. Two main pedestrian flows cross the street, giving rise to a large number of interactions. In this context, 190 pedestrian trajectories have been manually tracked, for a total number of 10200 observed positions. The collected data contain the pedestrian identifier, the time step and the image coordinates.



(a) (b)

Figure 5.4: Speed-time graphs for the same two pedestrians

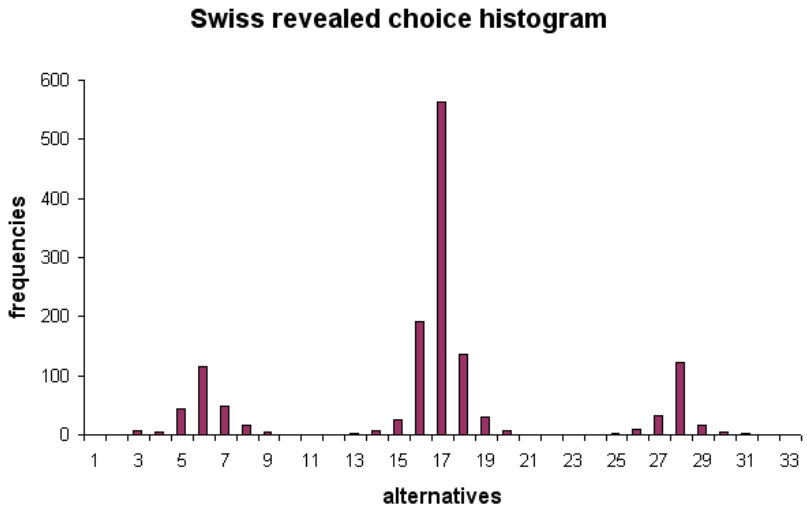


Figure 5.5: Revealed choices histogram

The mapping between the image plane and the walking plane is approximated by a 2D-affine transformation, whose parameters are learnt by linear regression (for more details we refer the reader to Teknomo, 2002). The reference system on the walking plane has the origin arbitrarily placed on the bottom left corner of the zebra crossing in Figure 5.6. The x axis represent the width of the crossing while y axis is the crossing length. We report in Figure 5.7 two example of trajectories and in Figure 5.8(a) and Figure 5.8(b) the related speed-time graphs. In Figure 5.9 we report the speed histogram and in Table 5.3 the speed statistics. Finally, the revealed choices histogram is reported in Figure 5.10, showing the three modes around the three alternatives along the central direction, as expected.

5.1.3 Data post-processing

The original Swiss dataset has been post-processed in order to generate the input data for the estimation process. At each step, the observed choice made by the current decision maker has been measured 3 steps forward in time, i.e. 0.9 seconds. As a consequence, the last four positions of each

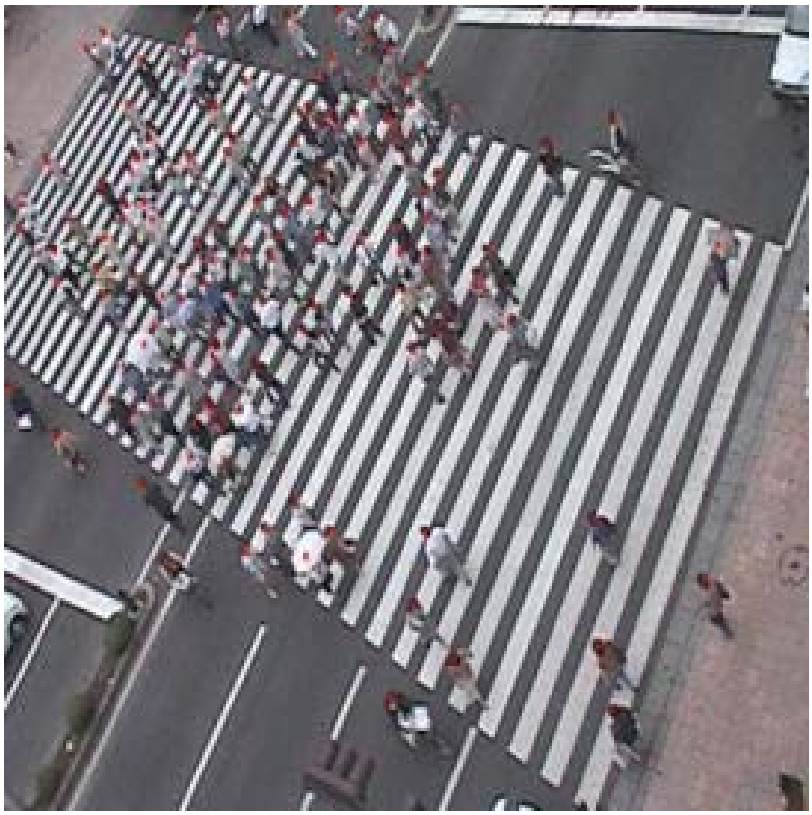


Figure 5.6: Japanese scenario

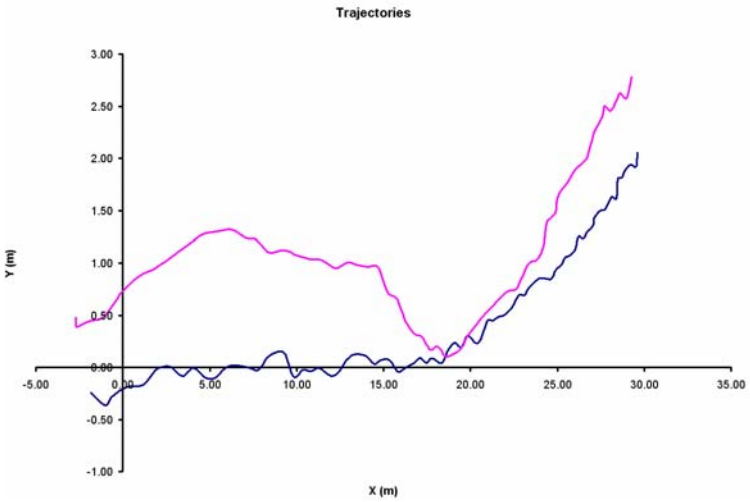
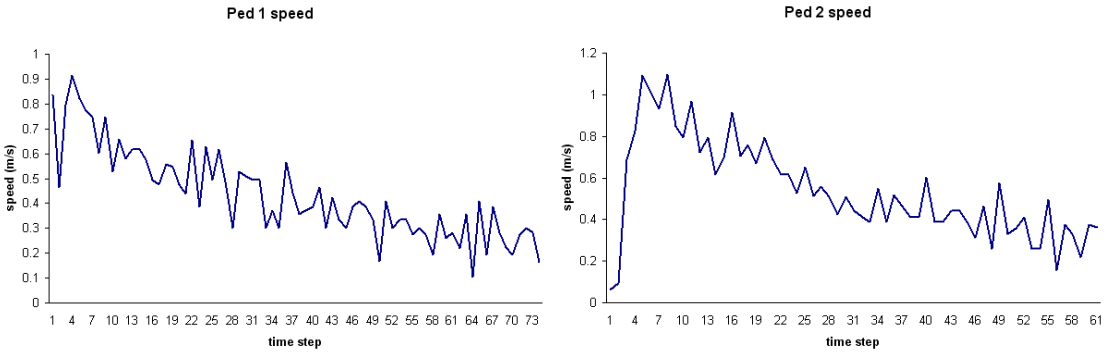


Figure 5.7: Examples of two manually tracked trajectories

trajectory are not used. Moreover, in both the datasets those observations corresponding to a static pedestrian ($v_n = 0$) and those corresponding to an observed choice out of the choice set have been discarded.



(a) (b)

Figure 5.8: Speed-time graphs for the same two pedestrians

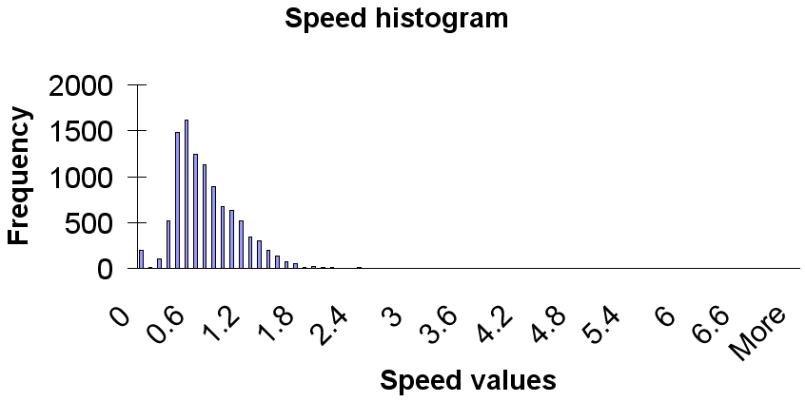


Figure 5.9: Speed histogram

We report in Table 5.4 and Table 5.5 the averaged values of the leader and collider availabilities (represented by the two indicator functions I_g^L and I_C defined above) defined as follows:

$$\begin{aligned} \bar{I}_g^L &= \frac{1}{N_S}\sum_{n=1}^{N_S} I_g^L \\ \bar{I}_C &= \frac{1}{N_J}\sum_{n=1}^{N_J} I_C \end{aligned} \tag{5.2}$$

where N_S and N_J are the two sample sizes.

5.2 Estimation results

In the previous section the datasets used for parameter estimation have been described. Collecting pedestrian trajectories through manual tracking from video sequences allows for the calibration of short range behavior models. We have seen in Chapter 4.2 different specifications. In this section the estimation results related to the different specifications are reported, using both the datasets,

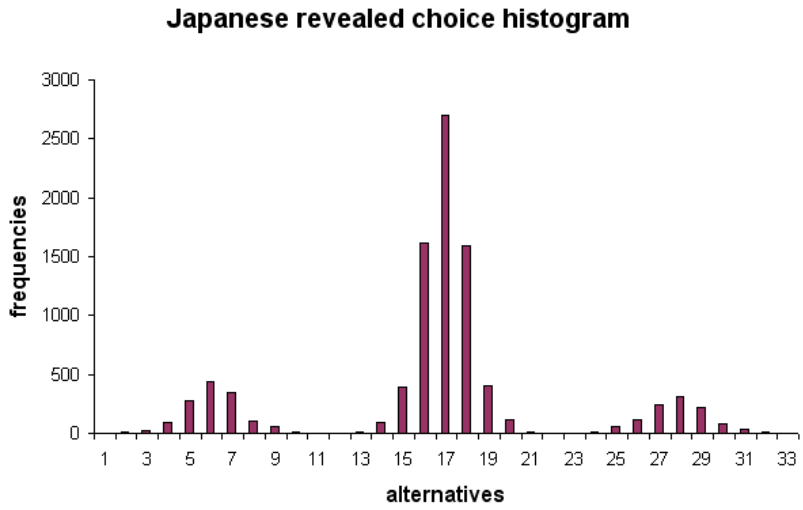


Figure 5.10: Revealed choices histogram

separately and pooled together. The aim is to present the results obtained at the different steps in the iterative modeling process. All the models have been estimated using the Biogeme package (Bierlaire, 2003), which is a freeware package for the estimation of a wide range of random utility models.

5.2.1 Step1

We discuss here the results related to themodel specification presented in Section 4.2. The parameters of the systematic utility functions are similar under different error structures, as shown in Table 5.7 and Table 5.8. It is a sign of reliability of the model specification, at least with respect to the dataset used for the estimation.
The $\beta_{direction}$ and $\beta_{destination}$ coefficients are both negative. This means that the utility of an alternative is going to decrease when its angular position is more decentralised with respect to the current direction and the final destination, respectively. This intuitive fact can be actually interpreted in a more interesting way. The tendency of an individual to keep her current direction and to move, if it is possible, toward her final destination can be seen as the evidence of decisions made at a higher level in the decision process. People decide their final destinations at a strategic level, their routes and paths at a tactical level (Hoogendoorn et al., 2002). These decisions are reflected on a short time horizon and are coherent with negative signs for the $\beta_{direction}$ and $\beta_{destination}$ coefficients. This conclusion is also accordant with other previous studies on the idea that individuals move through spaces along paths that minimize the angular displacements (see Turner, 2001 for more details).
The $\beta_{occupation}$ parameter has a negative sign, implying that individuals tend to prefer nearby spatial zones less crowded by other pedestrians. This follows our expectation but it requires a certain prudence in terms of generalizability. It is actually licit to expect a positive value of this parameter (based on its definition) in those situations where a lot of individuals are walking toward the same destination (crowded environments with lane formation). More insights are necessary to generalize such a behavior, working with larger datasets.
The β_{angle} parameter has not been estimated significantly different from zero. The collision avoidance behavior is an important part of the pedestrian walking behavior. The lack of a significant result in this case is due to a lack of pedestrian-pedestrian interactions in the Swiss dataset.

Finally, the acceleration model is also accordant with our expectations. The signs of the β_{acc} and β_{dec} are negative, revealing a preference for the constant speed alternatives. The absolute values of the two coefficients are strongly different. An early specification introduced a linear term. The implementation of such a model in the simulator has shown important limits. We were actually faced with undefined accelerations or sudden abrupt decelerations. The introduction of a non linear specification with the elasticity coefficients has solved the problem. The λ_{acc} and λ_{dec} parameters have the expected signs. $\lambda_{acc} > 0$ implies an actractiveness for accelerations which decreases when the speed increases. This trend is shown in Figure 5.11. The curve for the acceleration dummy shows that when the speed increases the $\tilde{\beta}_{\text{acc}}$ decreases, reaching an ideal negative asymptote in correspondence of a maximum speed. Similarly, $\lambda_{dec} < 0$ indicates that when the speed decreases the attractiveness for a further deceleration decreases, reaching an ideal negative asymptote for a speed value equal to zero. The current speed module is normalized with respect to the maximum value observed in the dataset, for numerical reasons. The correlation between alternatives is partially captured by the CNL formulation. Two over five nest coefficients are significantly different from one, the first capturing the correlation between the constant speed alternatives and the second the correlation between the not-central choices. The mixed NL formulation results, shown in Table 5.8, present a better explanatory power with a better final log-likelihood value (-2559.97 against -2579.25). Here, the correlation based on the speed is captured by two (over three) nest coefficients, significantly different from 1, for the constant speed and the decelerated alternatives, respectively. The correlation depending on direction is captured by the normal terms, for which six standard deviations have been estimated, significantly different from zero. It is interesting to note that those normal terms capturing the correlation between those alternatives belonging to far radial directions (from the central one), namely σ_1, σ_9 and σ_{10}, show larger standard deviation values. On the contrary, the same values for more central alternatives, namely σ_5, σ_7 and σ_8, are lower. A possible interpretation for this fact is that choices corresponding to central alternatives are easily interpretable based only on kinematic characteristics, while more stressed direction changes are more likely to be caused by other causes, not included in the model (depending for example on individual characteristics), leading to random parts of the corresponding utilities with a higher variance. It is an indication for a more detailed direction change model. In Figure 5.12 we report the values of the probability vs. speed, all the rest remaining constant, for the decelerated, constant speed and accelerated alternatives (28, 17 and 6, respectively) of the central direction. The speed value increases in the range $[0.1, 7.1]$ m/s, with a step of 0.1 m/s. The fixed values for the other attributes are reported in Table 5.6. The curves are as expected. The probability for an acceleration (deceleration) decreases (increases) when the speed increases. Note that the curve describing the constant speed alternative is higher with respect to the others, showing how actually individuals perceive variations in speed as costs.

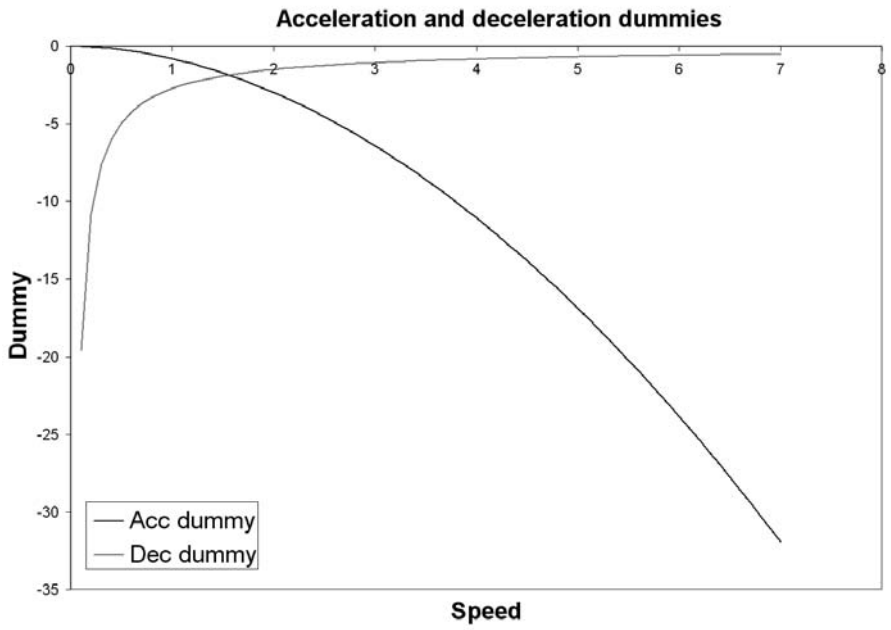


Figure 5.11: The parameters $\tilde{\beta}_{\text{acc}}$ and $\tilde{\beta}_{\text{dec}}$.

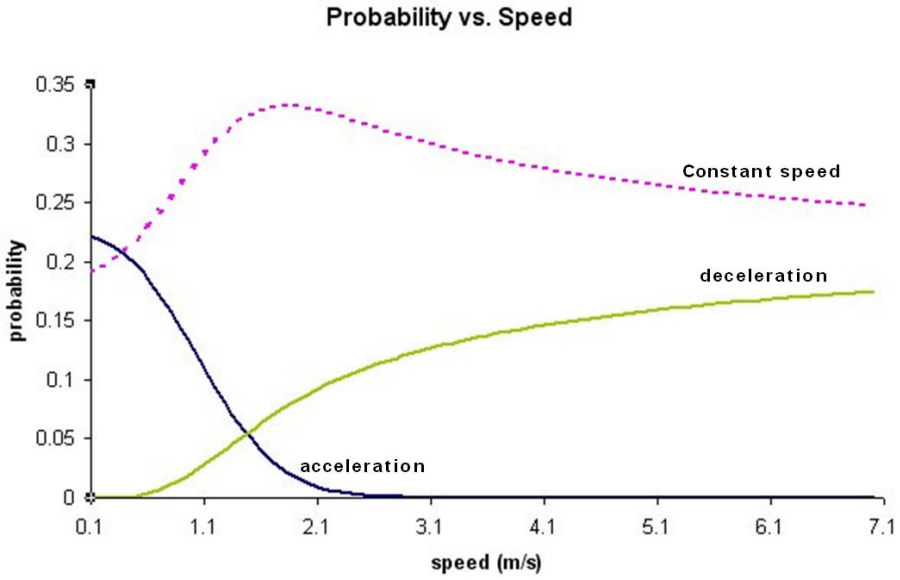


Figure 5.12: The probabilities as function of the speed modules, all the rest remaining constant in the utilities.

Pedestrian	Length	Av. speed module	Av. direction changes	Av. dest/dir angle
1	80	1.31	17.21	49.74
2	76	1.39	24.46	50.79
3	58	1.07	12.74	34.85
4	67	1.10	25.72	41.48
5	30	1.50	12.38	18.25
6	31	1.40	12.38	23.82
7	98	2.65	4.76	8.20
8	78	1.47	10.67	8.47
9	76	1.50	7.66	7.76
10	74	1.51	6.55	6.54
11	65	1.17	15.79	15.02
12	64	1.21	11.60	14.69
13	64	1.18	10.64	17.75
14	64	1.21	22.98	25.73
15	41	1.37	19.31	14.27
16	44	1.44	10.69	10.30
17	93	1.80	6.11	10.14
18	26	1.36	11.31	9.86
19	27	1.39	7.43	8.30
20	31	1.48	8.02	7.06
21	27	2.08	5.45	11.62
22	26	2.23	5.20	5.66
23	23	1.50	15.02	15.00
24	21	1.67	21.93	21.75
25	10	3.13	8.78	8.27
26	8	3.13	14.95	11.72
27	52	1.35	16.91	15.24
28	48	1.41	10.03	9.23
29	51	1.31	10.74	9.49
30	50	1.32	9.98	12.27
31	27	1.99	6.48	5.72
32	33	2.09	11.05	9.17
33	37	1.63	13.18	14.71
34	29	1.21	11.38	9.26
35	23	1.10	9.48	8.90
36	23	1.32	14.15	17.88

Table 5.1: Data statistics

Mean	1.577680303
Standard Error	0.018473153
Median	1.446021656
Mode	0.80741
Standard Deviation	0.697101509
Range	6.763794202
Minimum	0.244096667
Maximum	7.007890869

Table 5.2: Speed statistics

Mean	0.668242
Standard Error	0.003547
Median	0.58023
Mode	0
Standard Deviation	0.35826
Range	3.939786
Minimum	0
Maximum	3.939786

Table 5.3: Speed statistics

direction	**leader availability**		**collider availability**
	accelerated	**decelerated**	
1	0.004	0.004	0.145
2	0.006	0.013	0.117
3	0.004	0.014	0.148
4	0.002	0.017	0.142
5	0.003	0.021	0.150
6	0.001	0.012	0.152
7	0.001	0.015	0.116
8	0.004	0.016	0.111
9	0.002	0.016	0.136
10	0.002	0.006	0.104
11	0.0007	0.002	0.069

Table 5.4: Averaged leader and collider availabilities for the Swiss dataset

direction	leader availability		collider availability
	accelerated	decelerated	
1	0.07	0.11	0.45
2	0.09	0.13	0.47
3	0.07	0.12	0.47
4	0.06	0.10	0.44
5	0.09	0.14	0.45
6	0.10	0.16	0.44
7	0.08	0.13	0.45
8	0.05	0.10	0.44
9	0.05	0.10	0.48
10	0.05	0.12	0.49
11	0.05	0.10	0.47

Table 5.5: Averaged leader and collider availabilities for the Japanese dataset

alternative	direction	destination	occupation	angle
6	0	40.41084	0	0
17	0	40.41084	0	0
28	0	40.41084	0	0

Table 5.6: Attributes' values to compute the probabilities reported in Figure 5.12.

Variable name	Coefficient estimate	t test 0	t test 1
$\beta_{occupation}$	-1.7334	-2.4767	
β_{dir}	-0.0921	-12.3423	
β_{ddir}	-0.0615	-12.2974	
β_{acc}	-33.6222	-2.9011	
β_{dec}	-0.5036	-3.8286	
λ_{acc}	1.8322	9.0784	
λ_{dec}	-0.8650	-4.9753	
μ_{const}	1.7957	6.1643	2.7315
$\mu_{not\_central}$	1.2867	8.2960	1.8486

Sample size = 1424
Number of estimated parameters = 9
Init log-likelihood = -4979.03
Final log-likelihood = -2579.25
Likelihood ratio test = 4799.55
$\bar{\rho}^2 = 0.4802$

Table 5.7: CNL estimation results

Variable name	**Coefficient estimate**	***t* test 0**	***t* test 1**
$\beta_{occupation}$	-1.5031	-2.7973	
β_{dir}	-0.1170	-10.5978	
β_{ddir}	-0.0737	-9.3415	
β_{acc}	-32.7867	-4.4996	
β_{dec}	-0.4495	-5.4054	
λ_{acc}	1.7677	12.1902	
λ_{dec}	-0.8987	-8.7415	
σ_1	1.4875	2.3527	
σ_5	0.6850	3.7304	
σ_7	0.9284	5.3723	
σ_8	1.2338	5.9258	
σ_9	1.6298	5.7021	
σ_{10}	2.3415	4.8001	
μ_{const}	1.4067	10.5971	3.0636
μ_{dec}	1.3164	6.3777	1.5329

Sample size = 1424
Number of Halton draws = 2000
Number of estimated parameters = 15
Init log-likelihood = -4979.03
Final log-likelihood = -2559.97
Likelihood ratio test = 4838.11
$\bar{\rho}^2 = 0.4828$

Table 5.8: Mixed NL estimation results

5.2.2 Step2

We describe, from a qualitative point of view, those intermediate steps in the modeling process that triggered interesting problems and represented the motivations for the further development presented in Section 4.3, whose results are reported in the next section. The original specification described in Section 4.2 has been tested on the Japanese dataset, when it was available. Only the CNL specification was used for preliminary tests, giving the huge estimation time required by the mixed nested logit model. These intermediate results clearly show a lack in modeling pedestrian interactions. The occupation parameter in Eq. (4.1) is an aggregate attribute designed to capture the negative impact of crowding on the utility of the corresponding alternatives. The Japanese dataset, as already said in Section 5.1, refers to a situation where two main pedestrian flows walk in opposite directions. The $\beta_{occupation}$ parameter results positive when the original model is estimated on this dataset. This counter-intuitive sign shows that this parameter captures a mixed effect due to the presence of the two main pedestrian flows. In our original interpretation, this fact would have meant that people move towards crowded positions. Another counter-intuitive sign obtained in this test refers to the elasticity for decelerations. The λ_{dec} parameter results positive and significantly different from zero, leading to the conclusion that the attractiveness for a deceleration is going to increase when the speed decreases, against our intuition. These results have triggered the need for a more detailed pedestrian interaction modeling, leading to the leader follower and collision avoidance models presented in Section 4.2.

5.2.3 Step3

In this section we present the estimation results for the model specification given in Section 4.3. They are reported in Table 5.9.

We first shortly comment the results for those parameters related to the unconstrained models (toward destination, keep direction and free flow acceleration). This part of the model specification is practically the same as the previous model, specified in Section 4.2. The *toward destination* coefficients β_{ddir} and β_{ddist} have been estimated significantly different from zero, indicating that the assumption that destination distance and direction capture two different effects is supported by the data, being related to the 2D nature of the pedestrian movements. Their signs are negative, as expected, reflecting the tendency of individuals to move directly toward their final destination, through the shortest path. The interpretation for the coefficients of the *keep direction* and *free flow acceleration* models are the same as Section 5.2.1.

We now comment on the constrained models' parameters. For the *leader follower* behavior we note that in the case of an accelerated leader, 3 out of 4 parameters have been estimated significantly different from zero. The positive value for the α^L_{acc} multiplicative coefficient indicates that when a leader is present (or several potential leaders are present, so that the closest to the decision maker is considered), a leader's acceleration induces a correspondent acceleration on the decision maker. The negative sign for the distance exponential coefficient, ρ^L_{acc}, indicates that the influence of the leader on the decision maker acceleration behavior reduces when their relative distance increases, as expected. The positive sign for the speed exponential coefficient, γ^L_{acc}, shows that the utility of an acceleration increases with higher values of the relative leader speed, as expected. The same interpretation is given for the parameters correspondent to a decelerated leader. In this case we keep in the model also the exponential coefficient related to the direction, δ^L_{dec}, being the related t-test statistics equal to 1.642. Its negative sign is coherent with the leader follower behavior. It reflects the fact that in those cases where the leader's relative direction is higher, the influence of the leader on the decision maker is lower, resulting in a lower utility value for the decelerated alternatives.

Variable name	Coefficient estimate	t test 0	t test 1
β_{ddir}	-0.061	-19.066	
β_{ddist}	-1.614	-1.9749	
β_{dir}	-0.027	-11.342	
β_{acc}	-19.822	-5.847	
β_{dec}	-2.069	-2.651	
λ_{acc}	0.969	26.880	
α^L_{acc}	4.883	3.368	
ρ^L_{acc}	-0.657	-3.034	
γ^L_{acc}	0.869	9.877	
α^L_{dec}	4.061	6.278	
ρ^L_{dec}	-0.481	-4.280	
γ^L_{dec}	0.524	9.089	
δ^L_{dec}	-0.892	-1.642	
α_C	-0.0058	-4.639	
ρ_C	-0.313	6.748	
γ_C	0.781	3.318	
μ_{const}	1.597	32.413	12.119
$\mu_{not\_central}$	1.487	15.765	5.160
μ_{scale}	0.591	-	-8.565

Sample size = 10783
Number of estimated parameters = 19
Init log-likelihood = -78558.3
Final log-likelihood = -22572.7
Likelihood ratio test = 30260.3
$\bar{\rho}^2 = 0.4007$

Table 5.9: CNL estimation results for the specification of Chapter 4.3

For the estimation of the *collision avoidance* parameters, we fix the exponential coefficient related to the collider relative direction, δ_C, equal to 1 for numerical convenience. The other three free parameters have been estimated significantly different from zero. The multiplicative coefficient α_C is negative, as expected, indicating that those directions more likely to lead to a collision have a lower utility with respect to the central (current) direction, taken as the reference one for normalization purposes. The exponential coefficient related to the distance between the collider and the alternative, ρ_C, has a negative sign. It shows the fact that a more distant collider has a less negative impact on the alternative utility. Finally, the exponential coefficient related to the relative speed, γ_C, is positive, as expected. It captures the fact that faster colliders have a more negative impact on the utilities than slower individuals, being the possible collision perceived as an event happening before in time.
The correlation structure is captured by the cross nested specification. Three nest parameters have

been fixed to 1 while two are left free in the model, capturing the correlation between the constant speed and the not central alternatives. They have been estimated significantly different from 1, which guarantee the consistency with the random utility theory.
We finally comment on the heterogeneity in the dataset. We estimate the scale factor μ_{scale} for the Swiss data, which captures the variance of the associated error term.

We report in the following some graphics illustrating the marginal effects of the different variables for the constrained models. In Figure 5.13(a) and Figure 5.13(b) the effects of a stimulus variation (due to changes in the relative leader direction and speed) are shown. Figure 5.13(a) shows an accentuated variation in the leader acceleration term which decays quite quickly when varying its relative direction. Figure 5.13(b) shows the acceleration term (for a fixed decision maker speed equal to 3 m/s) when the leader speed is free to vary. As expected, higher acceleration values correspond to higher relative speed values, with a zero acceleration when the leader speed is equal to the decision maker speed, as expected. In Figure 5.14 the effect of variations in the sensitivity function (varying the leader distance) are reported. As expected, lower acceleration terms correspond to lower relative distance values. Finally, we report in Figure 5.15(a), Figure 5.15(b) and Figure 5.16 an example of the probability of a central deceleration (alternative 28) when varying the relative (decelerated) leader direction, speed and distance, respectively.

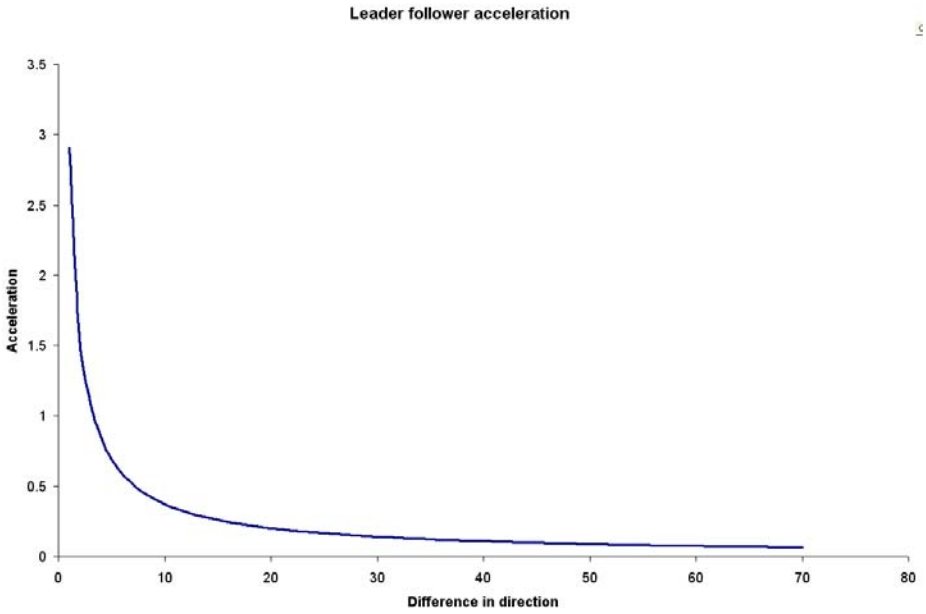
Similarly, in Figure 5.17(a) and Figure 5.17(b) we report the effects of variations in the stimulus term for the collision avoidance model. Figure 5.17(a) shows how for colliders coming from more frontal directions with respect to the decision maker direction (increasing relative direction), the collision term is reduced, reducing the alternative's utility. Figure 5.17(b) show how the collision term reduces for higher relative collider speed values. In Figure 5.18 the effects of changes in the sensitivity term are reported. It shows how farther colliders induced a lower negative effect on the utility, i.e. the collision term increases. Finally, we report in Figure 5.19(a), Figure 5.19(b) and Figure 5.20 an example of the probability of a central acceleration (alternative 6) when varying the relative collider direction, speed and distance, respectively.

5.3 Pedestrian simulator

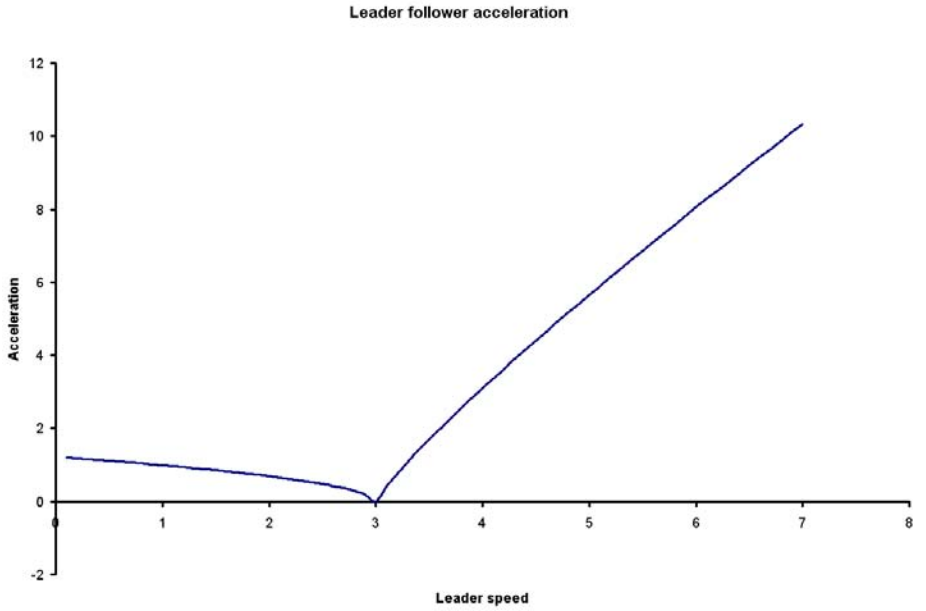
In order to validate our model, we needed to apply it so that we could verify if it represented realistic human walking behavior. We did indeed discover quite a few problems in early specifications of the model by using the simulator — problems that we probably would not have seen without its help. Initial versions of the model were instable with regard to speed: the pedestrians were either accelerating or decelerating to unreasonable speeds, which led us to introduce the speed elasticity parameters, which solved the problem by giving the pedestrians a stronger "will" to maintain a constant speed.

As shown in Figure 5.21 the addition of the simulator to our system creates a feedback loop that results in an enhanced model (Antonini et al., 2004b).

Other potential uses of simulation include capacity analysis, evacuation scenarios, incident scenarios, analysis of flow organization. Many developments exist in this area (see for instance Helbing et al., 2000, Kopp, 1999 and Gwynne et al., 1997) but to our knowledge, none of them uses a behavior model calibrated on actual data.



(a) Marginal effect of the relative leader direction



(b) Marginal effect of the relative leader speed

Figure 5.13: Effects of variations in the leader stimulus parameters

5.3.1 Design

There are essentially two approaches to simulation: *time-based* and *event-based*. In the time-based approach, the simulation proceeds in fixed time steps and all actors of the simulation are updated at each of these steps. In the event-based approach, events (e.g. collisions) are generated and inserted into a priority queue and are then executed in increasing time order. For now, we have chosen a

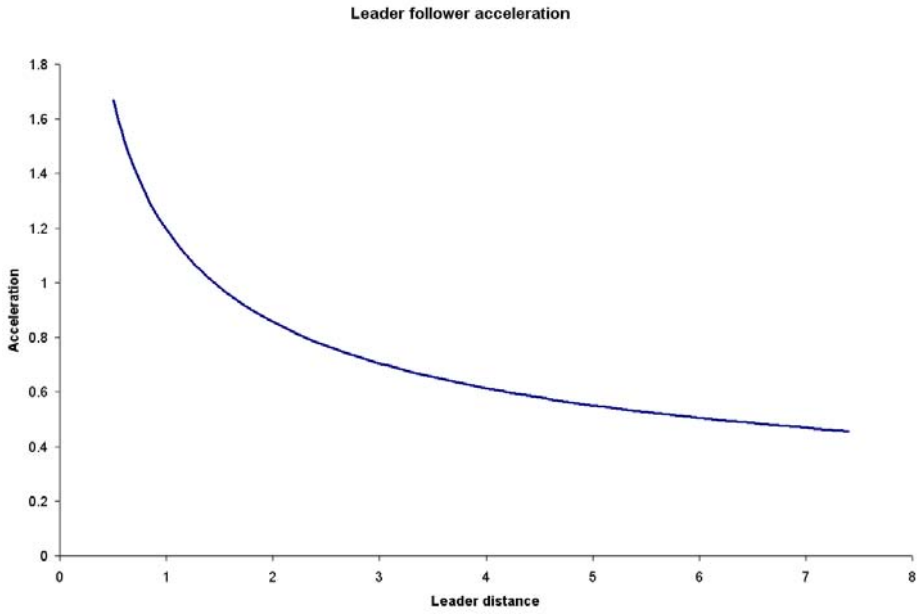


Figure 5.14: Effects of variations in the leader relative distance

time-based approach because the model is simpler, but we might move to an event-based approach later if the evolution of our model requires each footstep to be controlled precisely. We currently use a time step of $\Delta t = 0.9s$ in our simulations, consistently with the model assumption.

The simulator was developed using object oriented design techniques and written in standard C++. As input, it accepts a description of the cross nested logit model (preferred for the closed form of the choice probabilities) as specified in Section 4.2:

- the β_i coefficients in Eq. (4.10),
- the μ and μ_m coefficients in Eq. (3.9),
- the α_{jm} coefficients in Eq. (3.9).

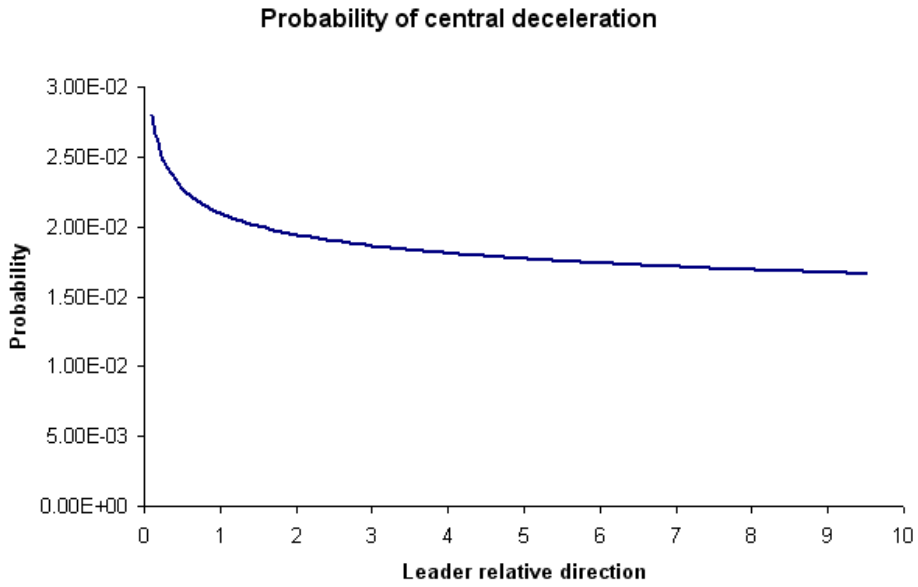
As output it produces images in a scene description language. We have successfully used *POVRay* for that purpose (`http://www.povray.org/`).

Following is a brief description of the algorithm inside our simulator:

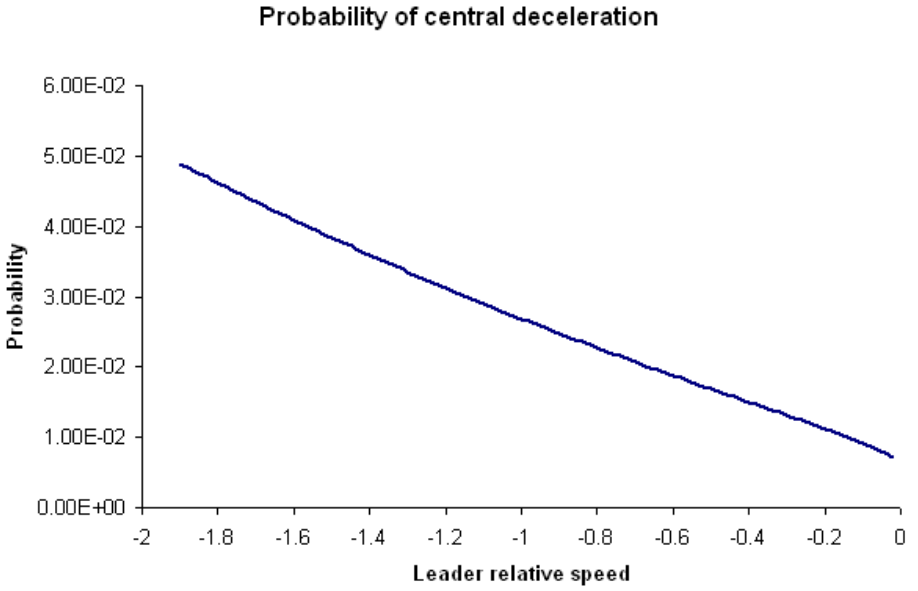
- **Initialization**

  The input to our simulator is a time-dependent origin-destination matrix, where each cell corresponds to an origin o, a destination d and a time interval ΔT, exactly like the OD matrices used for transportation applications. The cells contain the number of individuals departing from o, targeting d during the time interval ΔT.

  From the time-dependent OD matrix, we create a population of pedestrians. Each pedestrian can be associated with a list of characteristics which can be adapted to specific model requirements. This approach is consistent with the concept of demand simulation proposed by Antoniou et al. (1997) and Bierlaire et al. (2000). Our CNL model does not contain socio-economic characteristics. Also, we associate an itinerary with each pedestrian. An itinerary is defined as a sequence of intermediate targets, such that target k in the itinerary is visible from



(a) Probability of central deceleration as a function of the relative (decelerated) leader direction



(b) Probability of central deceleration as a function of the relative (decelerated) leader speed

Figure 5.15: Variations in probability as a function of the leader parameters

the position of target $k-1$, consistently with the network presentation presented in Bierlaire et al. (2003).

- **Moving decisions**

  First, new pedestrians are loaded in the system, with an initial speed corresponding to their desired speed, and an initial direction corresponding to the next target in their itinerary.

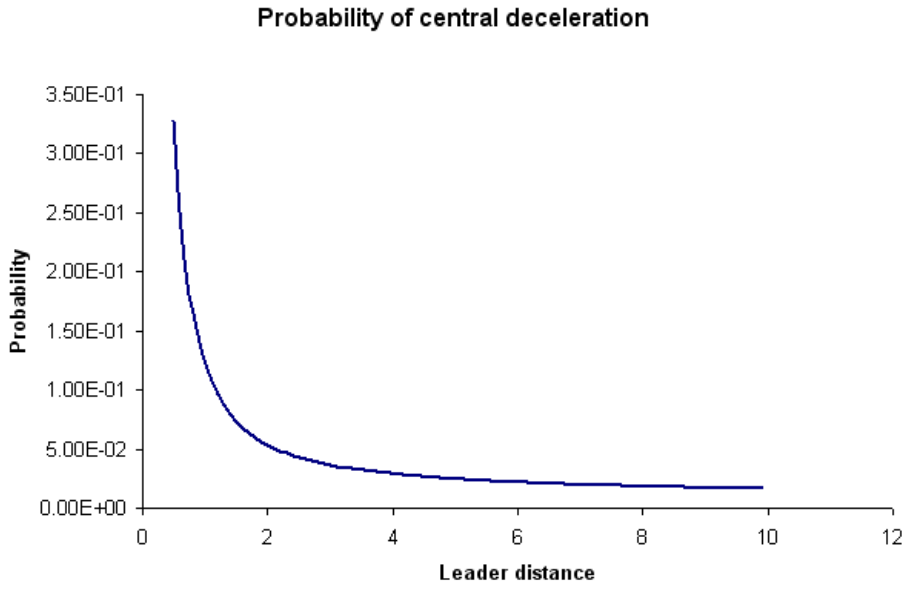
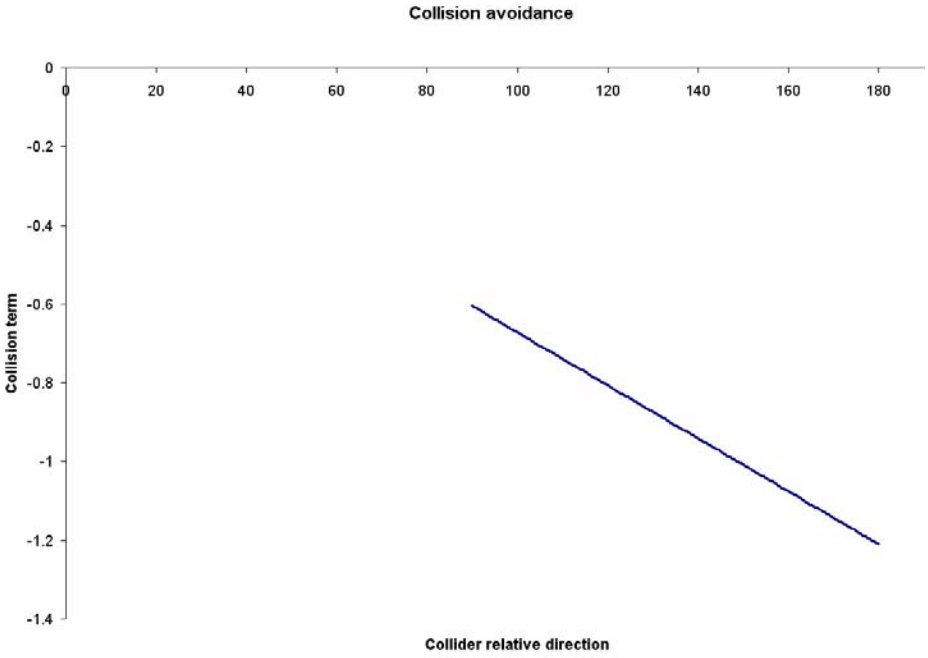


Figure 5.16: Probability of central deceleration as a function of the relative (decelerated) leader distance

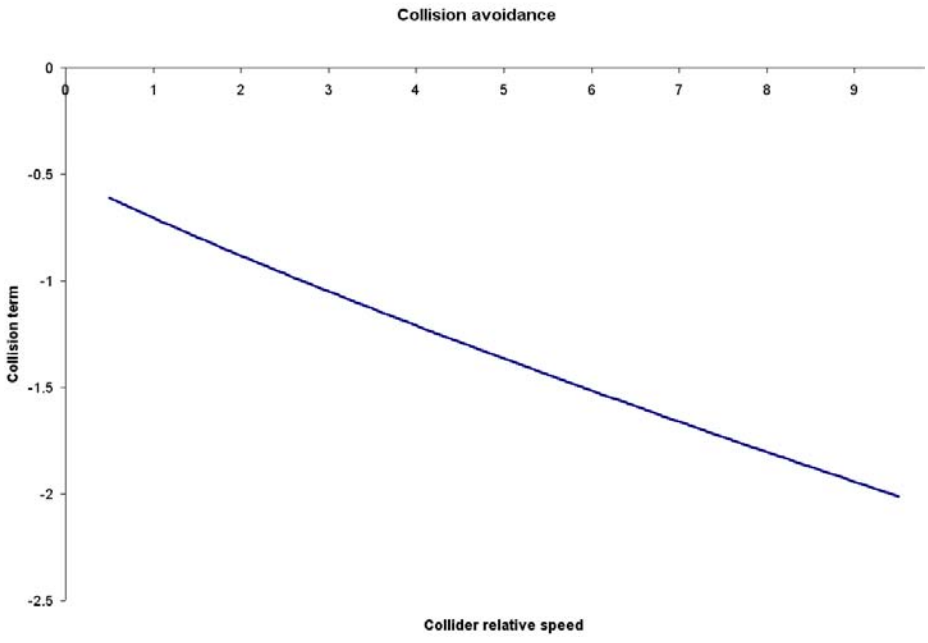
A *pre-step/step* logic is followed. This means that in the pre-step phase the next position for the agent is computed but the actual move is left to the step stage which is performed only in the case that the chosen next position results available. At each time step (Δt), the utility value of each possible move is calculated for each pedestrian. These values are then transformed into probabilities consistent with the discrete choice model. A move is declared infeasible (and the associated alternative unavailable) in the presence of a physical obstacle. We report in the following an example of one simulation step. Figure 5.22 represents an hypothetic simple scenario where 3 other pedestrians are present in the scene, at the indicated positions with the reported movement directions. For simplicity, we assume here that the current direction d coincides with the final destination direction D. An obstacle is present, leading to the unavailability of alternatives 1, 2, 3 and 4. We report in Table 5.10 the availability, systematic utility, probability values and the attributes' values for each alternative. Based on the computed probabilities, the pedestrian's choice is randomly selected according to these probabilities.

Once the chosen position is recognized to be available, the speed and direction of the individual are updated according to the formula $x_{i+1} = x_i + \Delta t\ v_i$, where x is the position, i the time step and v the speed.

Figure 5.23 shows a pedestrian as depicted by our simulator. Here the choice set is shown.



(a) Marginal effect of the relative collider direction



(b) Marginal effect of the relative collider speed

Figure 5.17: Effects of variations in the collider stimulus parameters

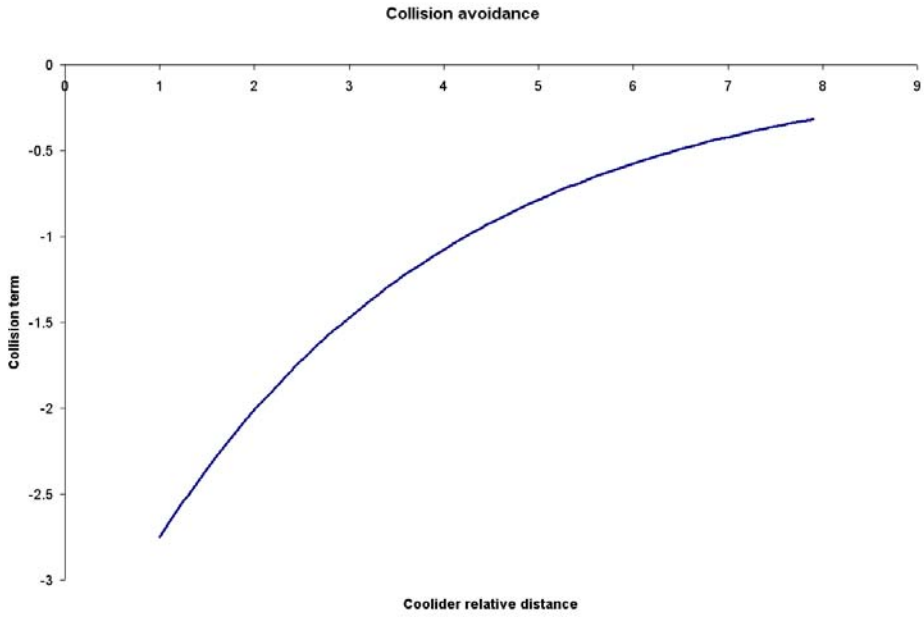
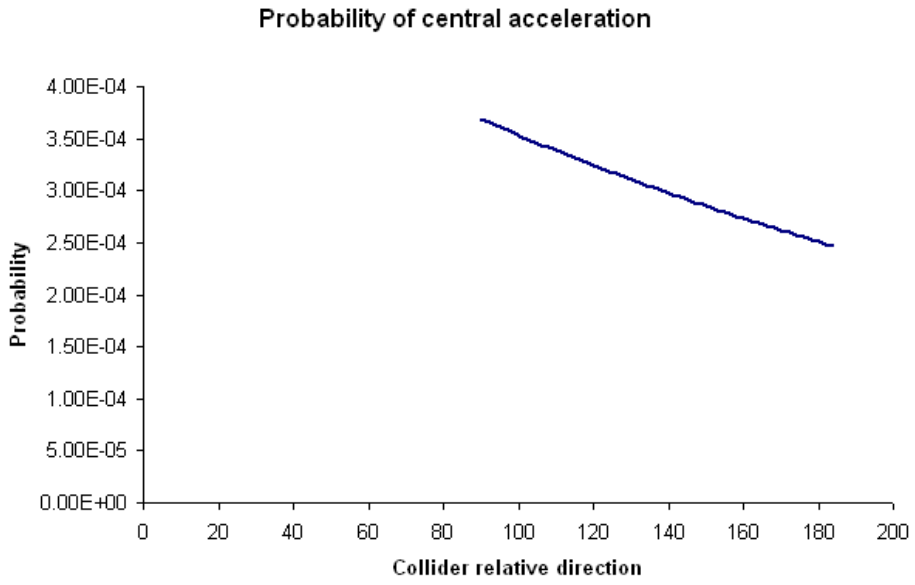
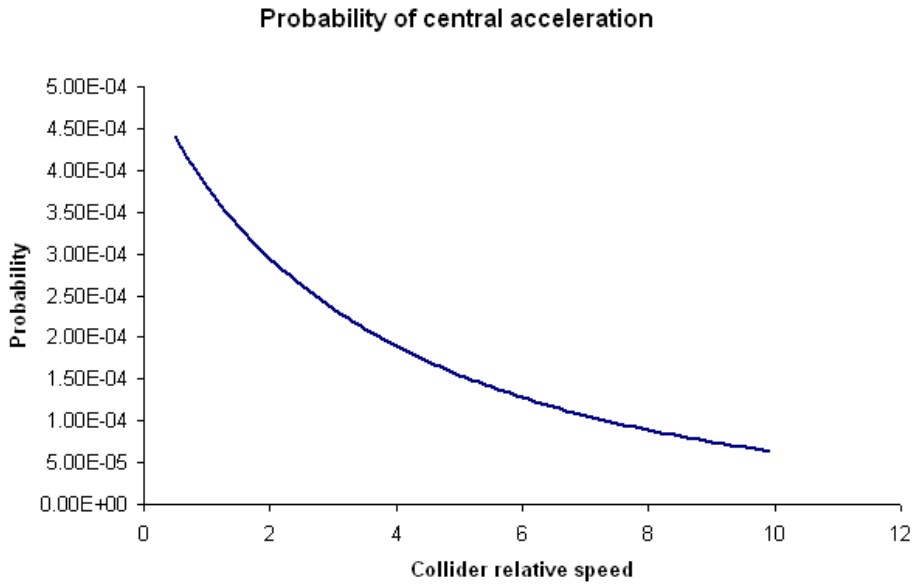


Figure 5.18: Effects of variations in the collider relative distance



(a) Probability of central acceleration as a function of the relative collider direction



(b) Probability of central acceleration as a function of the relative collider speed

Figure 5.19: Variations in probability as a function of the collider parameters

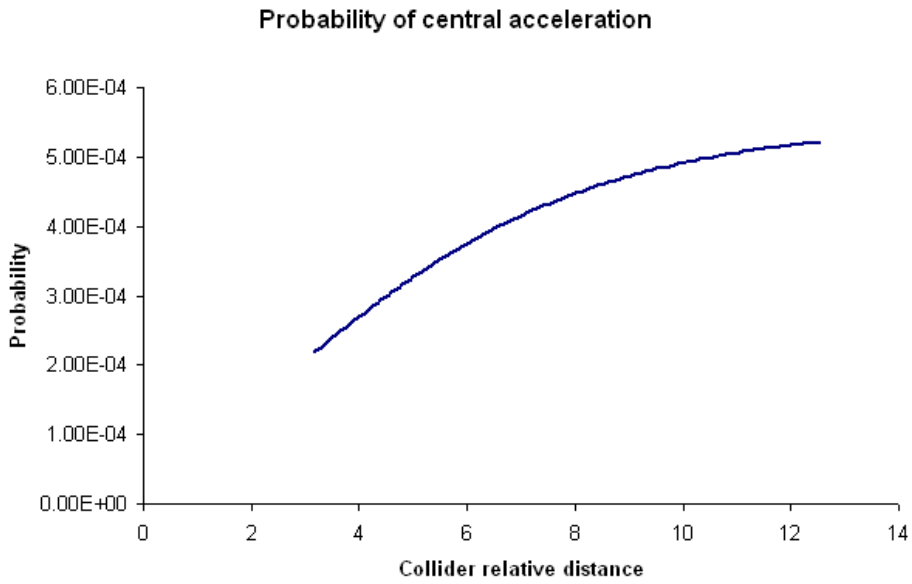


Figure 5.20: Probability of central acceleration as a function of the relative collider distance

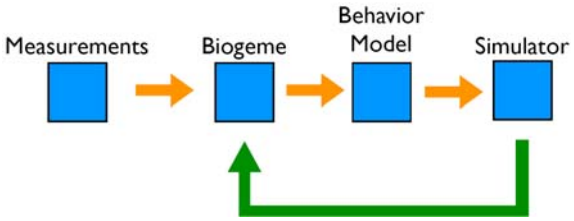


Figure 5.21: Model/simulator feedback

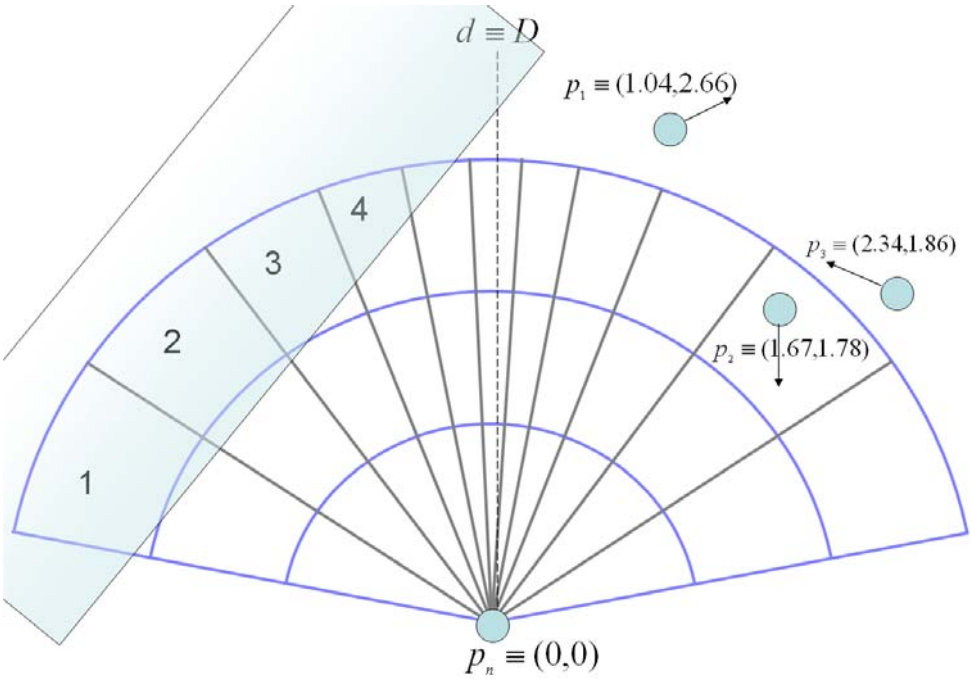


Figure 5.22: Example of a simulated scenario.

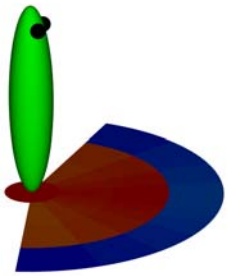


Figure 5.23: Pedestrian with choice set

alt	avail	v_n	direction	destination	occupation	V	prob
1	0	1.7	72.5	72.5	0	-	0
2	0	1.7	50	50	0	-	0
3	0	1.7	32.5	32.5	0	-	0
4	0	1.7	20	20	0	-	0
5	1	1.7	10	10	0	-4.05	7.64e-03
6	1	1.7	0	0	0	-2.51	5.09e-02
7	1	1.7	10	10	0	-4.05	7.64e-03
8	1	1.7	20	20	0.454	-6.37	6.46e-04
9	1	1.7	32.5	32.5	0	-7.51	1.98e-04
10	1	1.7	50	50	1.325	-12.49	1.22e-06
11	1	1.7	72.5	72.5	0	-13.65	3.8e-07
12	1	1.7	72.5	72.5	0	-11.14	2.37e-07
13	1	1.7	50	50	0	-7.68	2.05e-05
14	1	1.7	32.5	32.5	0	-4.99	6.8e-04
15	1	1.7	20	20	0	-3.07	8.8e-03
16	1	1.7	10	10	0	-1.54	7.37e-02
17	1	1.7	0	0	0	0	6.13e-01
18	1	1.7	10	10	0	-1.54	7.37e-02
19	1	1.7	20	20	0.217	-3.45	5.29e-03
20	1	1.7	32.5	32.5	0	-4.99	6.8e-04
21	1	1.7	50	50	0.639	-8.79	4.9e-06
22	1	1.7	72.5	72.5	0	-11.14	2.37e-07
23	1	1.7	72.5	72.5	0	-12.85	8.53e-07
24	1	1.7	50	50	0	-9.39	2.84e-05
25	1	1.7	32.5	32.5	0	-6.71	4.56e-04
26	1	1.7	20	20	0	-4.79	3.47e-03
27	1	1.7	10	10	0	-3.25	1.83e-02
28	1	1.7	0	0	0	-1.71	1.14e-01
29	1	1.7	10	10	0	-3.25	1.83e-02
30	1	1.7	20	20	0.108	-4.97	2.84e-03
31	1	1.7	32.5	32.5	0	-6.71	4.56e-04
32	1	1.7	50	50	0.296	-9.91	1.68e-05
33	1	1.7	72.5	72.5	0	-12.85	8.53e-07

Table 5.10: Simulation example. We report the attributes' values with the exception of the angle attribute, being the relative coefficient not significant.

5.3.2 Results

The videos generated by the simulator can be found at

`http://ltswww.epfl.ch/ltsftp/antonini/Simulator/`

Three different views can be generated by the simulator. In Figure 5.24 we illustrate the *normal view*. Pedestrians move on a real background (in front of a metro station in Lausanne) according to the calibrated CNL model. The *model view* is reported in Figure 5.25, where also the choice set is displayed. Finally, in Figure 5.26 we illustrate the *top view* obtained assuming to look at pedestrians from the top of the scene.

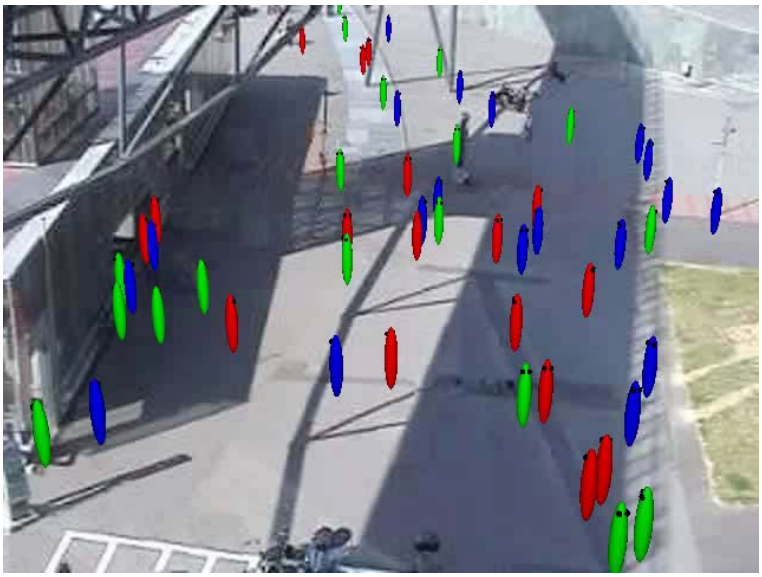


Figure 5.24: Normal view

5.4 Summary

In this chapter the data collection process, parameter estimation step along with a validation procedure have been discussed. Collecting data on pedestrian movements is a complex task. We used video sequences, manually tracking individuals over time and storing their spatial positions on the walking plane. These data are purely dynamic, allowing to extract speed and acceleration information. Two different samples, one collected in Switzerland and the other in Japan, are used separately and pooled together, in order to calibrate the proposed models. The maximum likelihood estimates

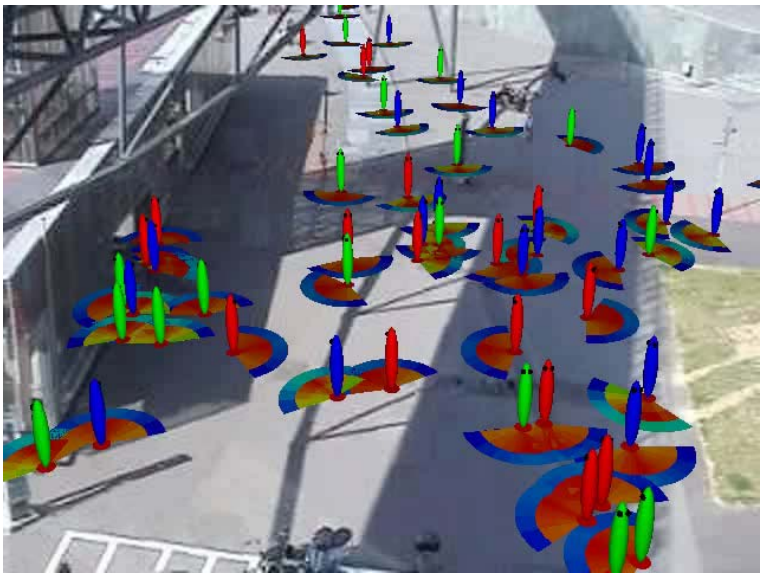


Figure 5.25: Model view

of the models' parameters are reported, with a discussion on their behavioral interpretation. Finally, a pedestrian simulator based on one of the proposed models, has been described. A simulation case is reported, showing how a it actually represents a precious tool for model validation.

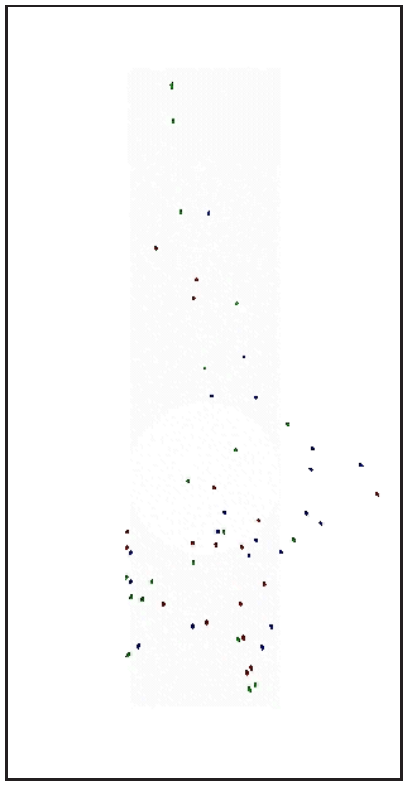


Figure 5.26: Top view

Part II

Application: Model-Based Pedestrian Detection, Tracking and Counting

Chapter 6

Pedestrian detection and tracking

In Part I the pedestrian behavioral model has been presented. In this second part an application is proposed, integrating the behavioral model for pedestrian tracking in video sequences. In this first chapter a review of the state of the art on detection and tracking methods is provided in Section 6.1. In Section 6.2 we describe our approach for pedestrian tracking and in Section 6.3 results are reported and discussed.

6.1 State of the art

Several different methods have been proposed in the last decade by researchers to address the object tracking problem. Among this multitude of approaches, different ways to classify them are possible. We first distinguish the problem of target detection from that of pure target tracking, adopting for the latter a two class partition, namely deterministic and probabilistic tracking methods.

6.1.1 Detection

Talking about tracking systems we are facing actually to a double problem. The first step consists in detecting a target of interest and only in a second time the detected target is tracked over time. Interesting surveys which include pedestrian detection are those of Gavrila (1999) and Lombardi (2001). This section is an extract from these works.
A first distinction in target detection methods is whether the steps of detection and target recognition are well separated or are not. Methods belonging to the first category usually perform an image segmentation step, where the foreground region is separated by the background, followed by a post-processing step where the foreground object is checked for the presence of the target of interest. Methods belonging to the second category, look for pedestrians in the whole image since the beginning and can be considered conceptually closer to the pattern recognition paradigm. A third (smaller) category can be identified with those methods where actually only foreground objects are detected in a *blind* way, assuming that all these image regions detected under a certain criterion are human subjects.

Motion detection

The information related to moving regions is certainly one of the most important cues to detect targets of interest. Some previous works use motion detection techniques (optical flow) to detect

individuals. Homogeneous regions under a certain criterion, as for example colour, shape or texture are checked to have similar values of optical flow. In this spirit, Bregler (1997) check for coherent motion of groups (blobs) of pixels of the same colour. Each pixel belongs to a given blob with a certain probability and each blob is classified using probabilistic models of motion. A space-time discretization is adopted by Polana and Nelson (1997), which assign to each image region (obtained simply overlapping a grid on the image) its averaged optical flow. Fourier analysis on the resulting feature vector is performed to detect periodic motion patterns. Shio and Sklansky (1991) use a motion field obtained by correlation techniques over successive frames and after some space-time smoothing followed by a quantization step of such a field, they identify motion regions with similar directions. Heisele et al. (1997) use groups of pixels obtained by clustering techniques as basic units for tracking. Both colour and spatial information is used to perform cluster analysis; the motivation for this is that adding spatial information makes clustering more stable than using only color information. The clusters are adapted over consecutive images by means of a k-means algorithm. Cutler and Davis (2000) use a "stabilized" change detection algorithm, with a time window τ. Followed by an appropriate thresholding, this operation provides a map of the pixels representing moving objects.

Background subtraction

Background subtraction represents probably the most used paradigm for foreground segmentation. Many approaches fall into the mixture-of-Gaussian framework. Wren et al. (1997) use a one single Gaussian distribution describing the colour. Foreground pixels are determined as those that lie some number of standard deviations from the background mean and are assigned to one between several Gaussian models corresponding to objects. The classification is based on the Mahalanobis metric. Stauffer and Grimson (2000) extend the statistical model, allowing for an adaptive mixture of Gaussians also for the background itself. Every pixel value is compared against the existing set of models at that location to find a match. The parameters for the matched model are updated based on a learning factor. If there is no match, the least-likely model is discarded and replaced by a new Gaussian with statistics initialized by the current pixel value. The models that account for some predefined fraction of the recent data are deemed "background" and the rest "foreground". Collins et al. (2000) augmented this approach with a second level of analysis to determine whether a pixel is due to a moving object, a stationary object, or an ambient illumination change. Oliver et al. (2000) use principal component analysis (PCA) on background eigenvectors resulting in a system analogue to the mean-variance map. A similar approach is used by Jabri et al. (2000) extending the idea to three mean-variance maps of colours, vertical and horizontal edges. Other statistics on the pixel intensity values (maximum and minimum, maximum of derivative over time and disparity) are used by Haritaoglu (1998). This approach works better than those based on simple mean-variance maps when pedestrian colours are closer to background colours. The drawback consists in too many false alarms for small objects with random movements. Paragios and Tziritas (1999) proposed an approach for finding moving objects in image sequences that further constrained the background/foreground map to be a Markov random field. Elgammal et al. (2002) used a non-parametric kernel density estimate for the intensity of the background and each foreground object. Pixel intensities that are unlikely based on the instantaneous density estimate are classified as foreground.

6.1.2 Pedestrian recognition

Shape-based models

Once a number of hypothetical targets are detected a decision has to be made about pedestrian/not-pedestrian. Classic approaches to pedestrian recognition are those based on shape models. The advantage of such methods is that they do not rely on any temporal information, so in principle they can recognize both moving and static targets. Papageorgiou and Poggio (1999) use a wavelet template to characterise a pedestrian shape and then scan the wavelet transform of the image to check for the presence of such a template. Wren et al. (1997) use a process guided by a 2-D contour shape analysis that attempts to identify various body parts using heuristics. Cai and Aggarwal (1996) describe a system with a simple head-trunk model to track humans across multiple cameras. Beymer and Konolige (1999) fit a simple shape on the candidates. They use an Ω model for the head and shoulders. This approach is very sensible to scale variation, so five different models are used, from coarse to fine resolution, according to the estimated distance of a subject. Fujiyoshi and Lipton (1998) use a skeletonization procedure to characterize the shape of a foreground object previously detected. For each object, they calculate first the centroid of the area and then the distances from the centroid to each border points. Local maxima of the distance function are taken as the external points of the skeleton. The authors suggest that the relative position of centroid and external points, and their rigidity, may be applied to recognition of different types of targets. For what concerns humans, they further confirm the analysis with gait detection. Kahn et al. (1996) use several cues including intensity, motion and edge. Many different object templates are defined and visual routines are provided to detect these in the images.

Gait recognition

Different approaches using temporal information relying on gait recognition. Human gait exhibits periodic characteristics and can be a very useful cue for walking pedestrians. The drawback of such methodology is that these methods do not allow for recognition of static pedestrians. Fourier-based methods are applied for periodicities. The candidate patterns varying over time are analyzed in frequency and classification techniques are used to select the human-like. Fujiyoshi and Lipton (1998), Cutler and Davis (2000), Polana and Nelson (1997) among the others use this approach. In Efros et al. (2003) the problem is recognizing actions from video taken from a distance, where the person appears only as a small patch. They compute a set of space-time motion descriptors on a stabilized figure-centric sequence, and match the descriptors to a database of pre-classified actions using nearest neighbour classification. Finally, Foster et al. (2003) use masking functions to measure area as a time varying signal from a sequence of silhouettes of a walking subject. Essentially, this combines the simplicity of a baseline area measure with the specificity of the selected (masked) area. The dynamic temporal signal is used as a signature for automatic gait recognition.

6.1.3 Pedestrian tracking

Once the target of interest is detected several approaches to track it over time can be applied.

Probabilistic tracking

Between the probabilistic methods for object tracking, special interest has been received in the last years by all those approaches based on a state space modeling for the target and looking at the tracking problem as a Bayesian filtering problem. We do not give here a detailed description of the Bayesian filtering theory, because out of the scope of this thesis. Such a statistical framework

is based on two stochastic equations. A dynamic equation describes how the state of the target evolves over time while a measurement equation updates the way how we can observe the state itself. Fall into this framework the classical Kalman filter, assuming linear dynamics and Gaussian error terms (for both the dynamic and measurement equation). In the last years, the non-linear non-Gaussian extension (particle filtering) has been of interest for many works in object tracking. We report here just a few of such (numerous) works (Isard and Blake, 1996; Kitagawa, 1996; Isard and Blake, 1998; Nummiaro et al., 2003; Nummiaro et al., 2002; Arulampalam et al., 2002; Thayananthan et al., 2003). Basically, the research in this direction has focussed on new state representations for the targets and different dynamic models, mostly based on the image plane. An interesting work in this direction is Thayananthan et al., 2003 where the state-space is partitioned using a tree-based representation and a 3D hand model is used as a prior. Different hand-poses are generated by the model and projected on the image plane. The posterior is represented using a piecewise constant distribution over the leaves of the tree. Thresholds on the posterior (on the different sub-trees) are used to converge efficiently towards the high-modes of the distribution. Maggio and Cavallaro (2005) propose a combination of particle filter and mean shift, and enhanced with a new adaptive state transition model. The proposed tracker first produces a smaller number of samples than Particle Filter and then shifts the samples toward a close local maximum using mean shift (Comaniciu et al., 2003, Han et al., 2004). The transition model predicts the state based on adaptive variances. A different probabilistic approach is that of Senior (2002). Here the author deals with appearance models from a probabilistic point of view. The target is represented using an RGB color model. The value given by the model in position x represents the appearance in that position while an associated probability mask gives the likelihood of the object being observed at that pixel. The tracking problem is formulated as a maximum likelihood problem. Sminchisescu and Triggs (2003) present an alternative approach to stochastic tracking using the Covariance Scaled Sampling (CSS) algorithm. CSS propagates a multi-modal prior, essentially a mixture of Gaussians, and locally optimizes the new estimates such that they correspond to local minima in the posterior. Minima are sought as optimization involves minimizing a cost function as opposed to maximizing pdf. During propagation, each Gaussian is sampled from according to the shape of the cost function, allowing sampling to be biased along the directions of most uncertainty. Mean-Shift tracking is another probabilistic algorithm that endeavours to maximize the correlation between two statistical distributions. The correlation, or similarity between two distributions is expressed as a measurement derived from the Bhattacharyya coefficient. Statistical distributions can be built using any characteristic discriminating to a particular object of interest. A general model might use color, or texture or include both. A very simple form of the image gradient, i.e., pixel intensity differences between nearest neighbours in both the x and y directions will be used as the random variables in the tracking distributions implemented here

Deterministic tracking

Deterministic algorithms assume that the human body position can be uniquely determined at each point in time, contrarily to probabilistic methods where a distribution is given. A first kind of deterministic methods refer to all those techniques where the tracking of a certain feature or target over time is based on the comparison of the content of each image with a sample template. These algorithms focus more on the *target representation* problem, dealing with the changes in the appearance of the target itself (Jurie and Dhome, 2001; Kaneko and Hori, 2003; DeCarlo and Metaxas, 2000; Terzopoulos et al., 1988). Luck et al. (2002) adopt a deterministic approach where they construct a visual hull using shape from silhouette methods and fit a body model to it using a physics based fitting mechanism.

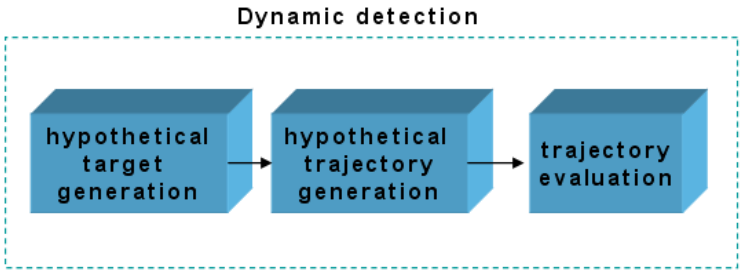


Figure 6.1: Overview of the dynamic detection approach

Most of the methods proposed in litterature for pedestrian tracking make use of the image-based information. The probabilistic framework provides a natural way to include any other source of knowledge, from different modalities to models for motion. In this context, the dynamic models used for tracking (usually in a state space modeling setup, from the first Kalman filter until the recent particle filtering) have been defined referring to a physical paradigm, where individuals are considered as particles subjected to forces. We propose a behavioral paradigm, where people move motivated by their destination, they follow other people in a crowd as a sub-optimal solution, avoid collisions with other individuals and perceive different discretizations of their surrounding space. In the next section we describe the integration of the model defined in Part I with a simple correlation-based tracking framework.

6.2 Model-based tracking

We do not make a clean distinction between detection and tracking. A simple *blind* initialization is performed, in order to reduce the computational complexity of this step. Both the deterministic and probabilistic approach to tracking are discussed (Antonini et al., 2005).

6.2.1 Dynamic detection

The idea is to recognize pedestrians on the base of their behavior over an *evaluation period*. In this spirit, our detection/recognition method can be classified as based on temporal information. The behavioral model presented in Part I is used to recognize those trajectories which are more likely to be generated by real pedestrians. A general overview of the dynamic detection method is illustrated in Figure 6.1.
Three main steps are required. First, an initialization step has to be performed in other to generate hypothetical pedestrians. Second, hypothetical pedestrian trajectories are generated from the initialization step. Third, the generated trajectories are evaluated by means of the calibrated model. The result of such an evaluation process is a filtering step, allowing to keep the most *human-like* trajectories. Such filtered trajectories represent the output of the dynamic detection system and correspond to the detected pedestrians. This approach to the detection problem differs from the state of the art basically for two main reasons

- the detection is based on the target's behavior rather than on the target's appearance;

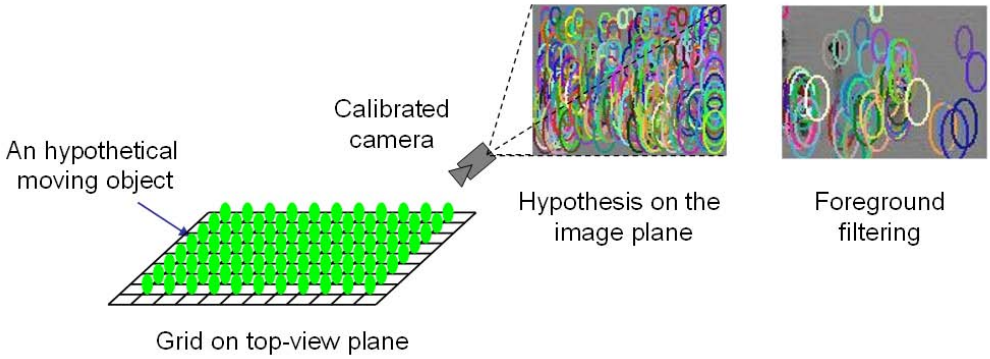


Figure 6.2: The initialization step

- tracking and detection are inter-operating steps. We need in fact a tracking method to build trajectories which will be evaluated over the evaluation period.

Hypothetical target generation

The initialization is performed in two steps. At first we place on the top-view a uniform rectangular lattice of points, at a resolution of 0.5m. Each of these points represents an hypothetical target to be detected and tracked. The topology and the resolution of the lattice can be tuned according to the *a priori* knowledge we have on the scene (exit and entry points, elevators, stairs etc...). The lattice structure is projected on the image plane by means of the calibrated camera, as shown in Figure 6.2.
Most of the video surveillance systems are actually equipped with fixed camera devices so it is relatively easy, in a real application context, to obtain the camera parameters. For this reason, our first operational constraint is the assumption to work with a monocular calibrated camera. It allows to define a unique correspondence between the image plane and the real walking plane, i.e. the top-view plane. There are two main reasons to work on the top-view plane. The target positions projected on the top-view does not suffer from occlusions, and the pedestrian behavioral model is defined on the real walking plane. As a consequence, the top-view represents the natural plane where image-related measures and behavioral constraints can be merged. Assuming the camera calibrated we know its parameters represented by the focal angle, the camera height, the angle with the horizon direction and the tilting angle around the vertical axis. So, given the pixel coordinates of an image point we can obtain unambiguously its projection on the top-view plane.
The second part of the initialization step consists in the use of a foreground mask, obtained by background subtraction. Any of the previous cited methods for background estimation can be used. In our case, for simplicity, we have an empty image of the scene. We can identify this initialization approach with a sub-sampling of the foreground region, driven by any specific knowledge on the layout of the scene, which is reflected in the topology and resolution of the grid. This approach represents a simple but effective way to initialize a tracking system, allowing at the same time for a low computational cost and fully automatic procedure.

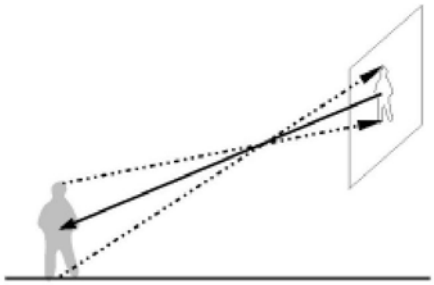


Figure 6.3: Automatic resizing of the target region.

Hypothetical trajectory generation

In this part of the algorithm we build step by step the hypothetical pedestrian trajectories which have to be evaluated. For each pair of consecutive frames we compute the displacement vector by maximisation of correlation. Given two images f and g of size MxN, the 2D discrete correlation between them is defined as:

$$C(x,y) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m,n) g(x+m, y+n) \tag{6.1}$$

for $x = 0, 1, ..., M-1$ and $y = 0, 1, ..., N-1$.
We aim to detect pedestrians looking at their behavior so we need information about their displacements rather than their appearance. Given an hypothetical pedestrian position $p \equiv (x, y)$ (on the image plane) and the corresponding image region $\hat{r}_t^p$ of size MxN centred around p at frame t, we compute the correlation $C(\hat{r}_t^p, r_{t+1}^p)$ between $\hat{r}_t^p$ and the corresponding region on the successive frame r_{t+1}^p. The maximum of the correlation gives the location $p_{max} \equiv (x_{max}, y_{max})$ of the best matching between the two image regions. The vector identified by the position of p_{max} with respect to p corresponds to the displacement vector of the current image region over the two frames. The interesting thing behind this well known method is that in two consecutive frames a human being can cover a limited distance, so it is reasonable to think that the searching region, used for correlation computation, contains the real target position. As it will be explained in the next section, we apply behavioral constraints to the trajectories generated by the motion vectors, projecting each of them on the top-view and stored in a buffer of length E_p while the p_{max} position in the successive frame is used to make a resizing of the region of interest. Assuming an averaged height of the human beings equal to 1.70 m, we obtain an automatic resize of the hypothetical target region on the image (see figure Figure 6.3), avoiding the use of complex deformation models to take into account changes in appearance given by the perspective, and introducing a negligible approximation error.
This step is repeated for each hypothetical target present on the current image and for E_p frames. In order to be able to detect any new target coming into the scene later in time, we need to periodically refresh the top-view grid at the image border. We illustrate the refresh grid in figure Figure 6.4. The refresh period, R_p, is assumed to be $R_p < E_p$.

Trajectory evaluation

The evaluation of the hypothetical trajectories is made in two steps.

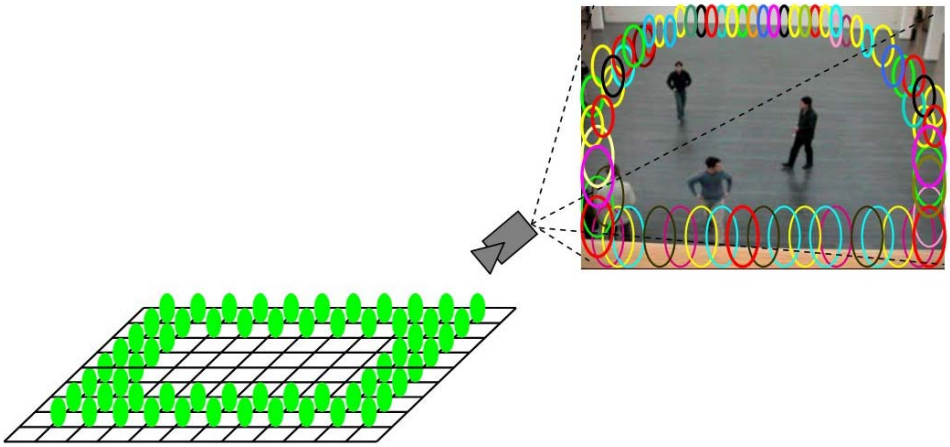


Figure 6.4: The refresh grid.

Pre-filtering We start to evaluate each of the E_p displacement vectors for each trajectory using simple distance and angular thresholds on the top view. This step is necessary to filter out the obvious outliers. In fact, many hypothetical targets placed for example on the shadows, can arise from noise (represented by some *impurity* on the foreground) or come simply from correlation errors. The goal of this preliminary step is to avoid the behavioral model computation for such outliers. In this stage we verify the projected displacements d_t^n and direction changes $\Delta\theta_t^n$ of the hypothetical moving objects, defined as:

$$d_t^n = \mathbf{p}_t^n - \mathbf{p}_{t-1}^n, \tag{6.2}$$

$$\Delta\theta_t^n = \theta_t^n - \theta_{t-1}^n \tag{6.3}$$

where $\mathbf{p}_t^n$ represents the position of the visual tracker n at time t, and θ_t^n represents the direction of the displacement between the positions $\mathbf{p_t^n}$ and $\mathbf{p}_{t-1}^n$. Following the idea to filter targets based on their dynamic, we give a cumulative *score* to a pedestrian trajectory over an evaluation period E_p. We implement these ideas with simple thresholds on the projected displacement vectors defining:

$$I_t = \begin{cases} 0 & \text{if } \|d_t^n\| \leq t_d \text{ and } \|\Delta\theta_t^n\| \leq t_\theta \\ -1 & \text{otherwise} \end{cases}$$

where t_d and t_θ are the thresholds on one-step distance and direction change. Studies on pedestrian dynamics (Schreckenberg and Sharma, 2002) show that the average speed value (in free-flow conditions) of a pedestrian is about 1.34 m/s. Our frame rate is 10 fps so we fix t_d to 13 cm. With analogous considerations we set t_θ to 120 degrees. The I_t is the one-step score given to a trajectory. We assign at each tracker an *activation* value representing the starting score and we decrement it at each 'bad' step. The final score for a tracker, S, is evaluated assuming a certain tolerance ξ to *bad steps* along the trajectory. We keep the tracker if the following condition is satisfied:

$$S = \frac{1}{T}\sum_{t=1}^{T} I_t \geq S_{inf} \tag{6.4}$$

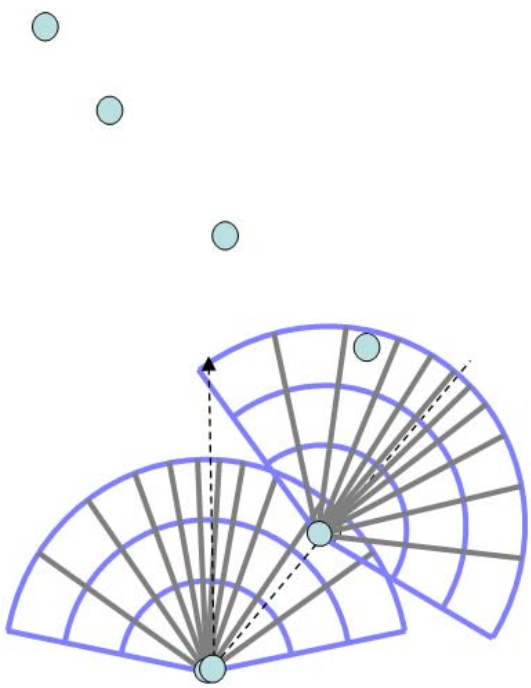


Figure 6.5: Each step in the collected hypothetical trajectory correspond to and alternative in the choice set of the calibrated model and has associated a probability value.

where S_{inf} represents the minimum score for a good trajectory. In our experiments we use $\xi = \frac{activation - S_{inf}}{activation} \geq 0.3$, which means a margin of 30% (we tollerate 3 'bad' steps over 10). The important parameters that have to be tuned are the *activation* and the evaluation period E_p.

Filtering The pre-filtered trajectories are the input for the behavioral filter. Each step done by an hypothetic pedestrian along his trajectory represents a choice made by the individual and it is characterised by a probability value given by the model, as illustrated in Figure 6.5.
We detect pedestrians giving a mark to the trajectory k based on the cumulative value of probabilities:

$$M_k = \frac{\sum_{l=1, j \in C_n}^{L} P_{jl}}{\sum_{l'=1, j' \in C_n}^{L} \max_{j' \in C_n} (P_{j'l'})} \geq th \tag{6.5}$$

where $j, j^{'} \in C_n$ are the alternative indexes in the choice set C_n, l and $l^{'}$ refer to the single step, L is the number of steps in the trajectory k, P_{jl} is the step probability as given by the discrete choice model and $\max_{j' \in C_n}(P_{j'l'})$ is the highest probability value associated with the most likely position at each step. The *th* value has to be fixed. This thresholding operation measures how much the collected score is far from the maximum probability score.

We want to underline the fact that the output model probabilities at each time step (which will give us the associated scores) are computed knowing the other pedestrians position, speed and direction but assuming those variables as stationary at the time of decision making for the current individual.

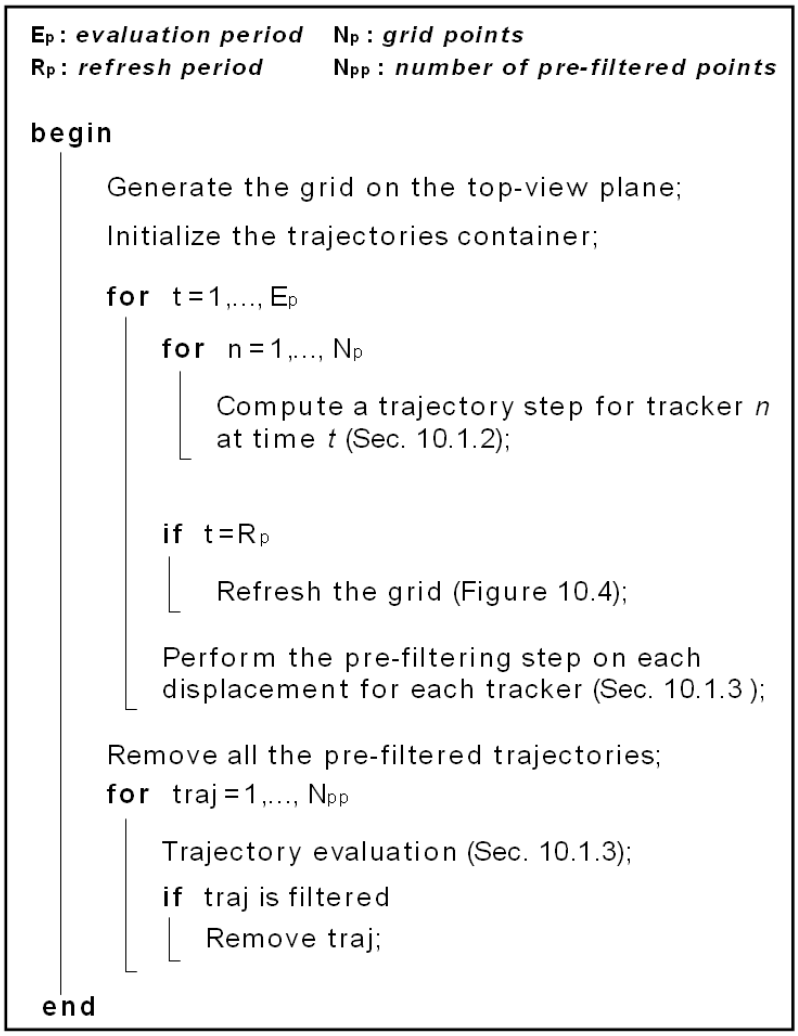


Figure 6.6: The dynamic detection algorithm

So, the interaction terms with the other pedestrians are implicit in the utility expressions (and hence are *mapped* into the probability values), defining how people perceive different positions as a function of both individual parameters and parameters related to the presence of the other pedestrians. The algorithm for dynamic detection is summarized in Figure 6.6.

6.2.2 Tracking

Deterministic tracking

One interpretation of the tracking problem is to treat it as an object detection made in each frame. Following this idea, the first implementation of the tracker is made repeating the dynamic detection algorithm.

Probabilistic tracking

In the first approach the behavioral model has used basically as a filter. Given a set of trajectories, we keep the most *human-like*. In the probabilistic implementation we adopt a Bayesian framework

$$P(M|D) \propto P(D|M) \cdot P(M) \tag{6.6}$$

to build trajectories frame after frame, once the dynamic detection has been performed. The implementation of the Bayes formula is made identifying the $P(M)$ term with the model probabilities and the likelihood term $P(D|M)$ with the following normalised correlation function:

$$NC^i_{t,t+1}(h,k) = \frac{C^i_{t,t+1}(h,k)}{\sum_l \sum_m C^i_{t,t+1}(l,m)} \tag{6.7}$$

where $C^i_{t,t+1}(h,k)$ represents (h,k)-element of the correlation matrix between $\hat{r}^i_t$ and r^i_{t+1} for the i-th pedestrian and the denominator is the sum of all the elements of the matrix. This normalisation implies that the probability of finding the pedestrian i in a certain position inside the r^i_{t+1} region is proportional to the corresponding correlation value.*

6.3 Results

6.3.1 Test sequences

Two outdoor sequences have been used to test the algorithm. The first sequence, *Flon*, has been recorded in front of the Flon metro station in Lausanne, 2002. The complexity of the scene is very high. Several pedestrians are present at the same time, moving in a scene with a highly cluttered background, many shadows areas and a deep (more than 100 meters) visual field. As a consequence, the target deformation induced by the perspective field is huge, leading to pedestrians of only a few pixels in the upper part of the images. Partial and total occlusions between pedestrians and group of pedestrians are frequent in this sequence. We report in Figure 6.7 a sketch of the Flon scene. Five different entry/exit points are present. From A to D they represent access points to the public square in front of the metro station. E are the elevators. As a consequence of this urban layout, five main bi-directional flows of pedestrian movement are present (arrows from **a** to **e** in Figure 6.7).

The second sequence, *Monaco*, has been recorded in Monaco by the Maia Research Institute. † This sequence is less complex than the Flon one. The visual field is restricted and only one bi-directional pedestrian flow is present. In the test sequence one individual is walking from left to right while 8 people walk in the opposite direction.

As already said in Chapter 6.2, for both the sequences a monocular static camera has been used, and an approximated camera calibration step has been performed, in order to use the top-view projection of the scene.

*This formulation contains the following approximation: the model is always propagated on a *maximum a posteriori* estimation of the posterior distribution. In this way, multi-modalities of the posterior are not taken into account

†http://www.maia-institute.org/

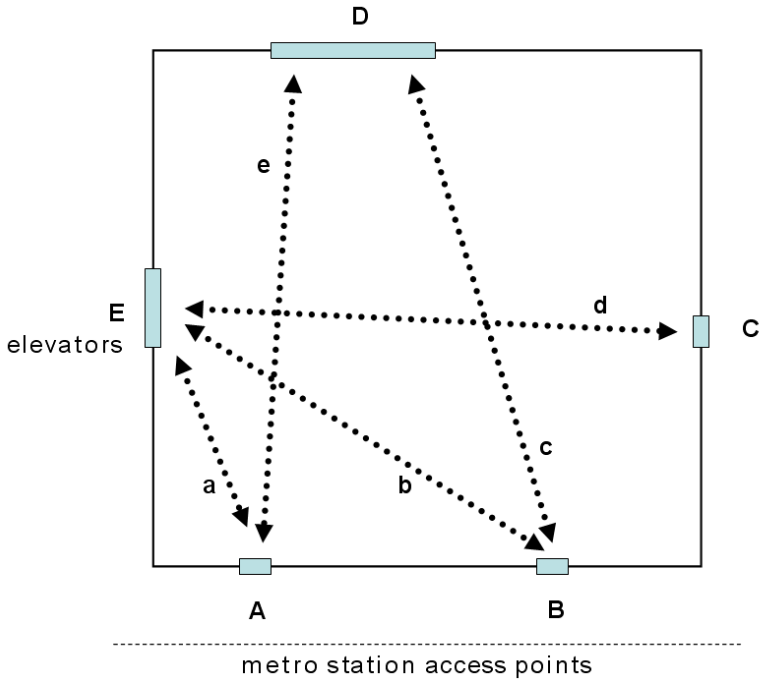


Figure 6.7: A sketch of the Flon scene. Five entry/exit zones are present. The arrows represent the main directions of pedestrian flow

6.3.2 Dynamic detection

In Figure 6.8 and Figure 6.9 we plot the number of filtered trackers as a function of the trajectory length (i.e. the evaluation time E_p in Figure 6.6), for different resolutions of the top-view grid for the two test sequences. It is interesting to note that the number of moving regions, associated with the moving points, presents a good stability. It means that we have a good degree of independence from the choice of the grid resolution and the evaluation time. In Figure 6.10 the three families of curves correspond to three different evaluation periods E_p. For each couple of curves, the dotted one represents the number of trackers after the pre-filtering, while the solid one refers to the output of the filtering step. We note that for low activation values (lower starting score of trackers), most of the filtering task is performed by the pre-filtering module. The DCM does not perform in this case any further filtering (the two curves overlap). Increasing the activation value (for example to avoid to loose at once good trackers), we see that a consistent further selection is done by the behavioral filter, as expected.

The temporal resolution of the model is of the order of 1 second. The idea is to observe people for a few walking steps, before to decide about pedestrian/not-pedestrian. The evaluation period and the activation parameters are logically correlated, in the sense that one refers to *how long* do we want to judge a trajectory and the other refers to *how permissive* we are in the evaluation process.

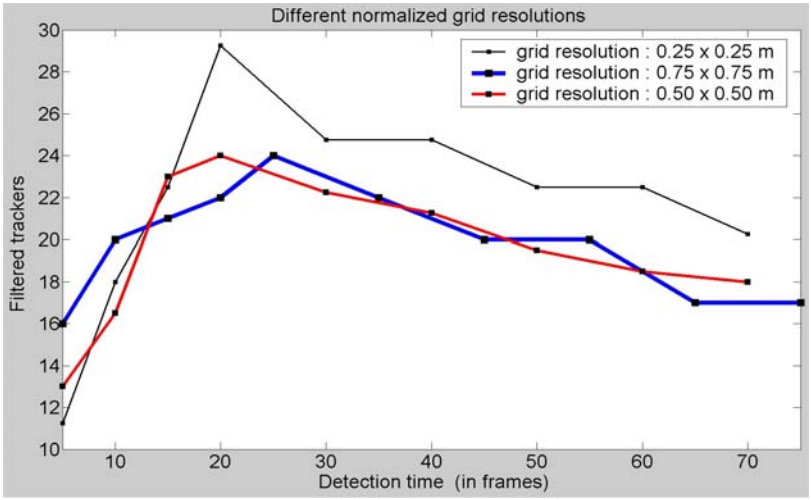


Figure 6.8: The number of filtered trackers for the Flon sequence, as a function of the evaluation time T for three different grid resolutions

In Figure 6.11 we report an example of dynamic detection for the *Flon* sequence. On the left-side image it is shown how the correlation process makes the trajectories' shape *noisy*. After the application of the behavioral filter most of the noise is removed, obtaining the human-like trajectories. The model filters the data at a trajectory level, and not at a single step level, so it is possible that good points are rejected if they are part of a trajectory not accepted by the filter. There is a tradeoff between the need to evaluate a whole trajectory, in line with the dynamic detection idea, and the need to avoid too strict threshold values.
In Figure 6.12 detection results at different frames are shown, for both the Flon and Monaco sequences. Considering the scene's complexity, the obtained detection rate shows that the behavioral filtering discriminates between noisy and *human like* trajectories, with a reasonable error rate. The

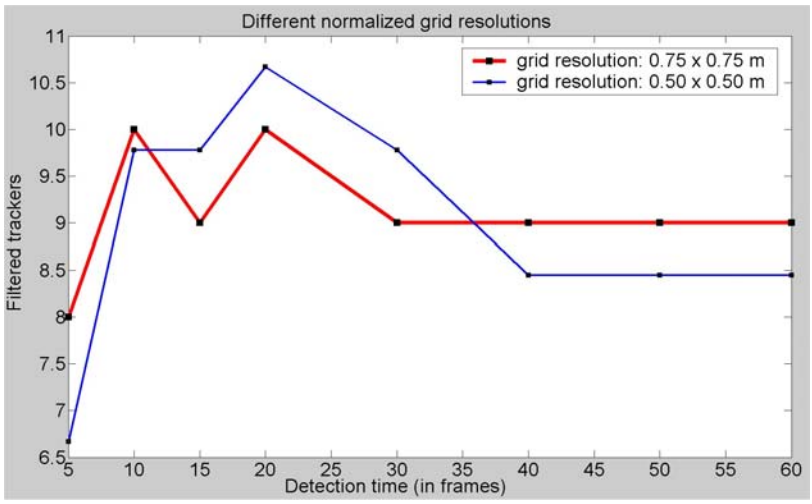


Figure 6.9: The same graphic as the previous figure for the Monaco video sequence, using two different grid resolutions.

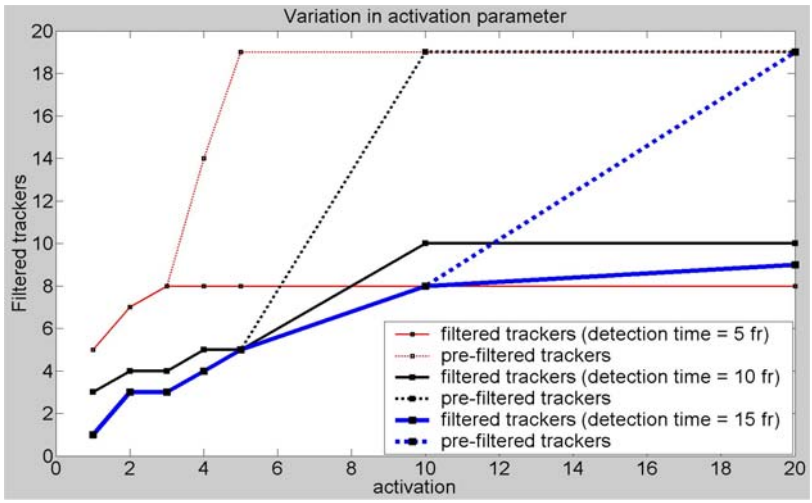
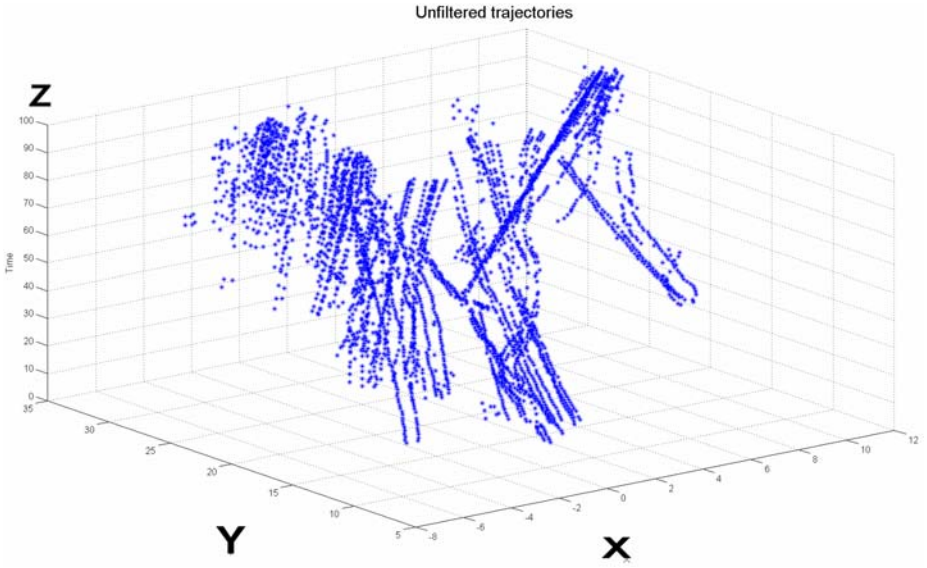


Figure 6.10: The variation of the filtered trackers as a function of the activation parameter. It shows the different roles of pre-filtering and filtering stages.

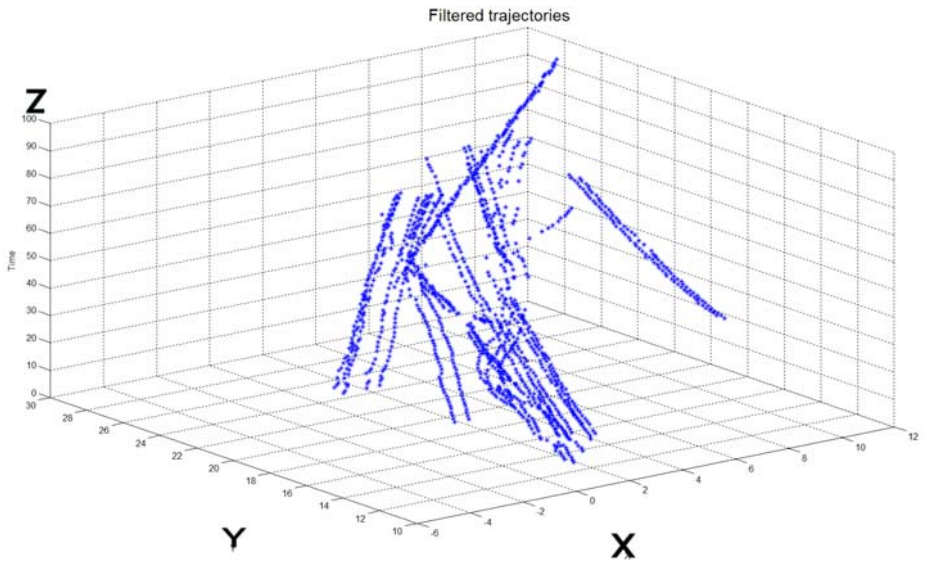
drawback of the system is the false alarm rate. This is due to the initialization step. Actually, the use of a grid on the top-view plane allows to avoid complex target detection procedures, at the price of an over-estimation of the real number of the targets. Multiple trackers placed on the same human body (or its shadow) give rise to multiple accepted trajectories. This problem represents the motivation for the work presented in Chapter 7.

6.3.3 Deterministic tracking

We report in Figure 6.13 and Figure 6.14 the results obtained from successive detection cycles, for both the Flon and Monaco sequences. The figures related to the Flon sequence show the results on both the image and top-view planes. The color of the tracks is related to the tracker identity.



(a) Un-filtered trajectories



(b) Filtered trajectories

Figure 6.11: Behavioral filtering. The x and y axes refer to the walking plane (in meters). The zero point on the x-axes corresponds to the camera position. The z axes represents the number of frames.

The tracking results show good performances of our system, given the complexity of the analysed sequences. It remains the problem of the target over-estimation. Some failures arise also from the pre-filtering step. In Figure 6.13 we have a positive detection at frame 65 (the yellow bounding box on the right) which disappears after a few number of frames. This failure is due to the pre-filtering step, showing the disadvantages of using fixed thresholds.

6.3.4 Probabilistic tracking

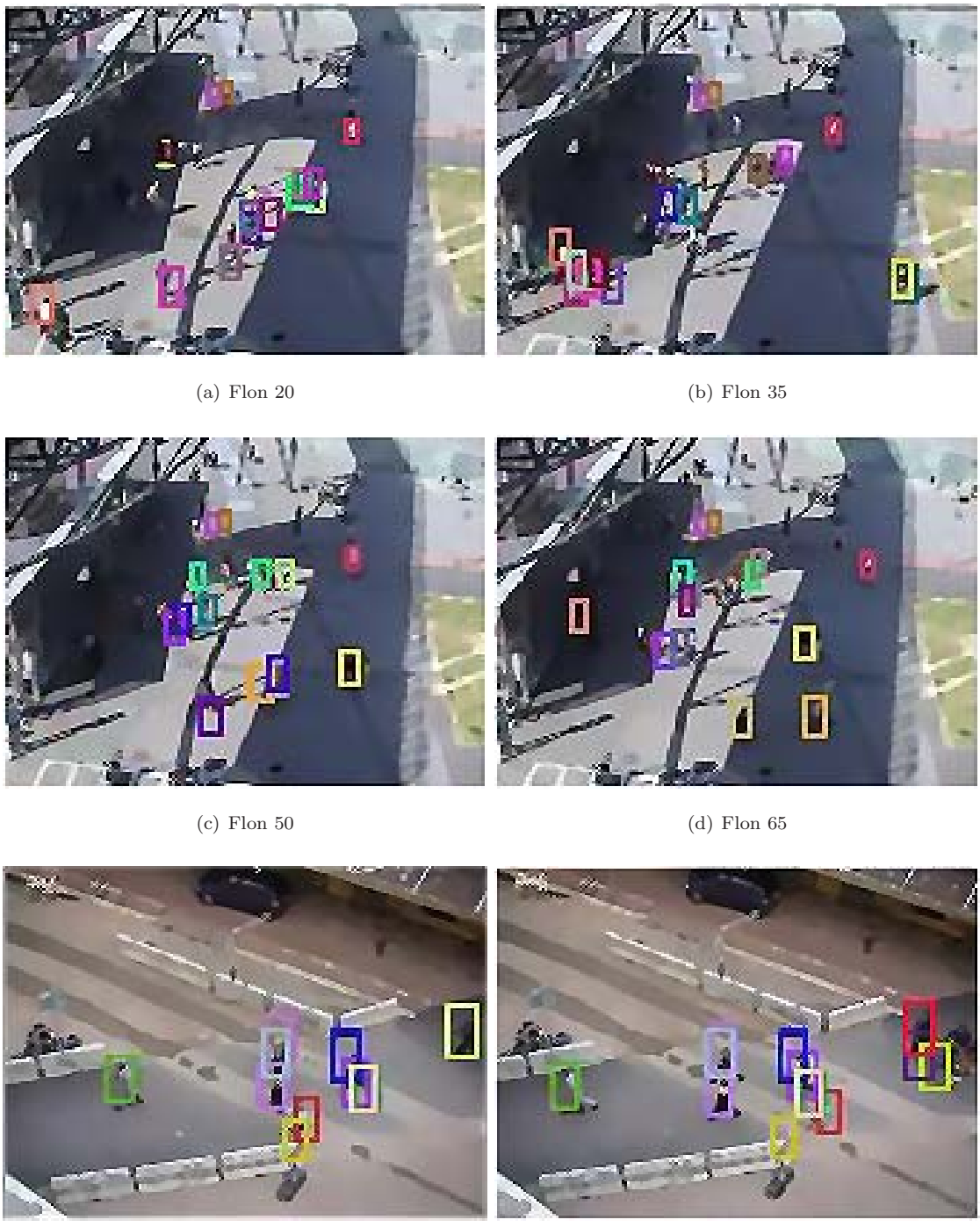
In Figure 6.15 and Figure 6.16 we show the results of the probabilistic implementation of the tracker. We compare the results obtained with a pure correlation-based tracker to those obtained integrating the model. In the first example the blue tracker (pure correlation tracker) does not follow the target in the dark zone. This problem disappears with the model-based tracker. Similarly, in the second example an application of the model to the case of tracker's jumps is illustrated.

The importance of objective evaluation protocols for tracking systems, especially for automatic video surveillance systems, has been underlined in the last years in both the research and the industrial communities. The IEEE international workshop on "performance evaluation of tracking and surveillance" (PETS), and the recent related special session at the 6th international workshop on image analysis for multimedia interactive services (WIAMIS 2005), show such a growing interest. The result of such numerous research efforts consists in a series of different (and sometimes confusing) methods and criteria, showing that at the moment we are far from an objective, recognized method of evaluation for tracking systems, depending strongly on the application field. Moreover, the generalizability of such a method often depends on different constraints on the recording devices (calibrated camera or self-calibrated systems, static or moving devices, monocular or stereo devices, different resolutions). For these reasons a validation step based on visual inspection has been adopted. The full video sequences showing the results are available at the web page

http://ltswww.epfl.ch/ltsftp/antonini/

6.4 Summary

The principal aim in the described approach is to investigate the integration of a behavioral model for pedestrian dynamics into a detection/tracking system. The image processing part has been kept simple, because of the preliminary nature of the work. The first two blocks in Figure 6.1 can be implemented, in principle, by any of the existing methods described in Section 6.1 for target detection and object tracking. The unique constraint is represented by the time resolution of the target search step, which has to be compatible with the proposed pedestrian behavioral model. Nevertheless, important conclusions have been reached. First, the use of a behavioral approach is not only reliable, but can also be extended maintaining the same mathematical framework, to higher levels of the individual decision process, which become fundamental for activity recognition and scene analysis. Second, dynamic detection is a powerful concept that integrates both detection (in the strict sense) and tracking together. The system has been tested on medium-high complex sequences, with cluttered background and multiple targets and occlusions, showing very encouraging results.

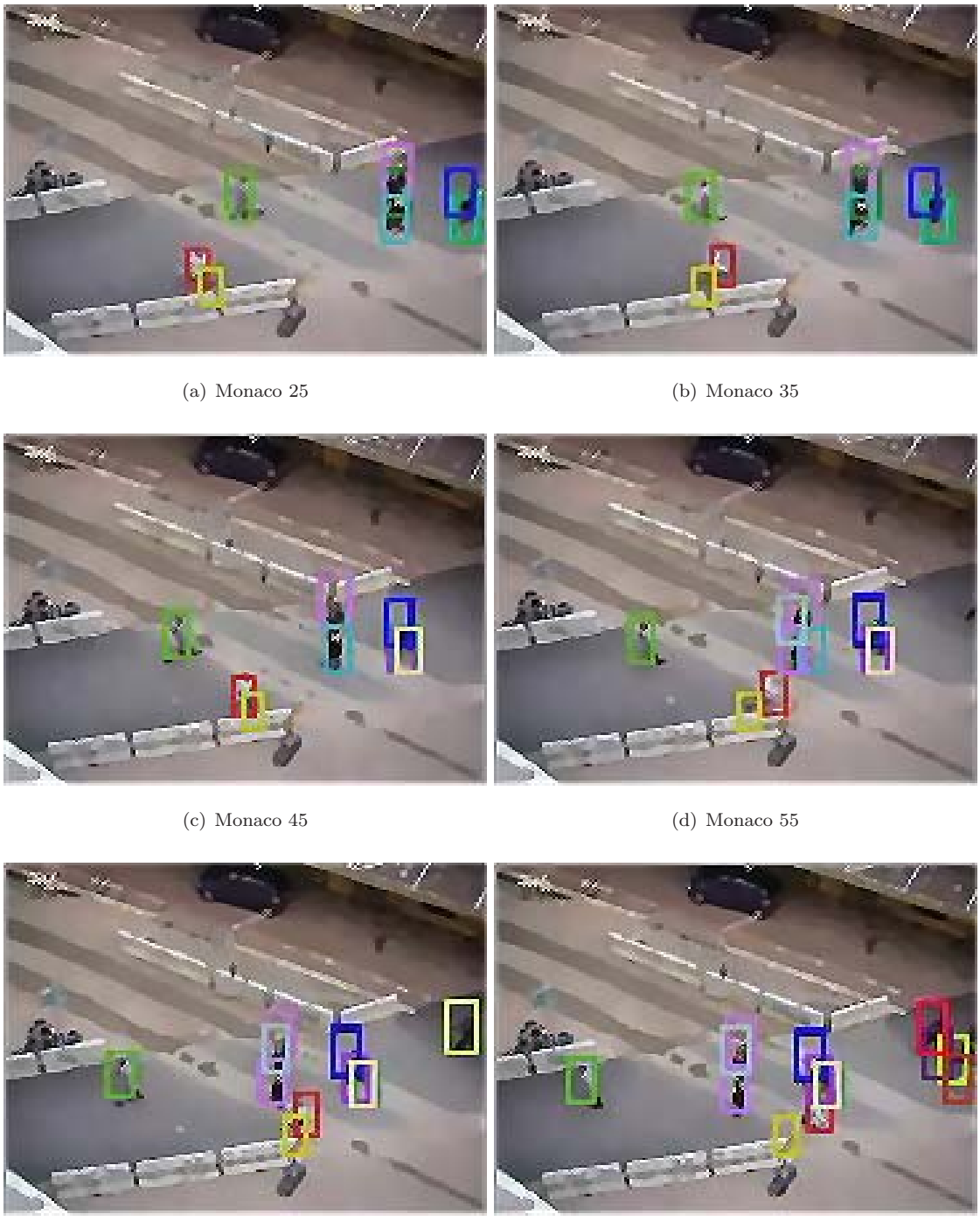


(a) Flon 20 (b) Flon 35

(c) Flon 50 (d) Flon 65

(e) Monaco 65 (f) Monaco 80

Figure 6.12: Some dynamic detection results

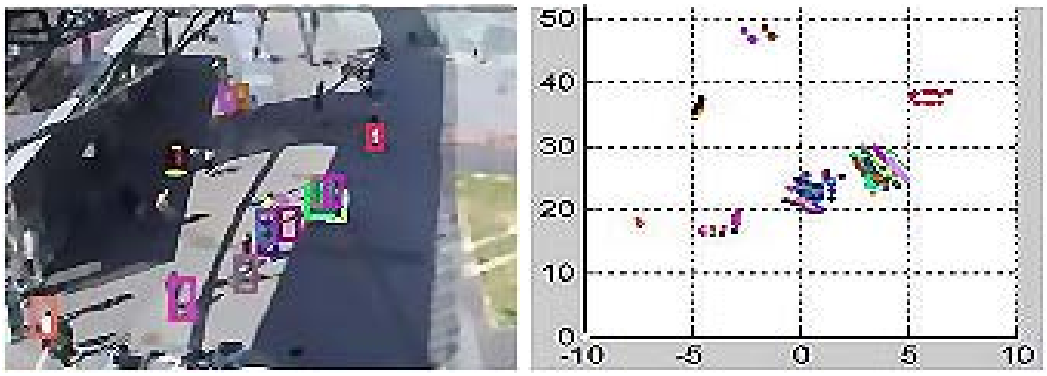


(a) Monaco 25 (b) Monaco 35

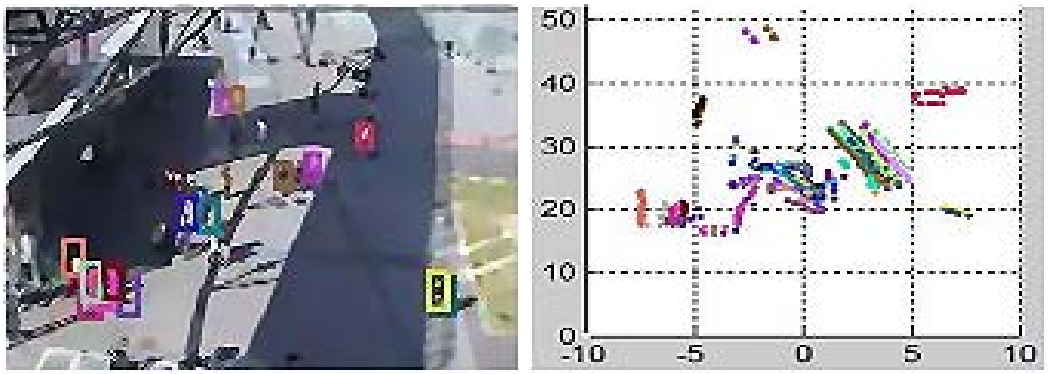
(c) Monaco 45 (d) Monaco 55

(e) Monaco 65 (f) Monaco 75

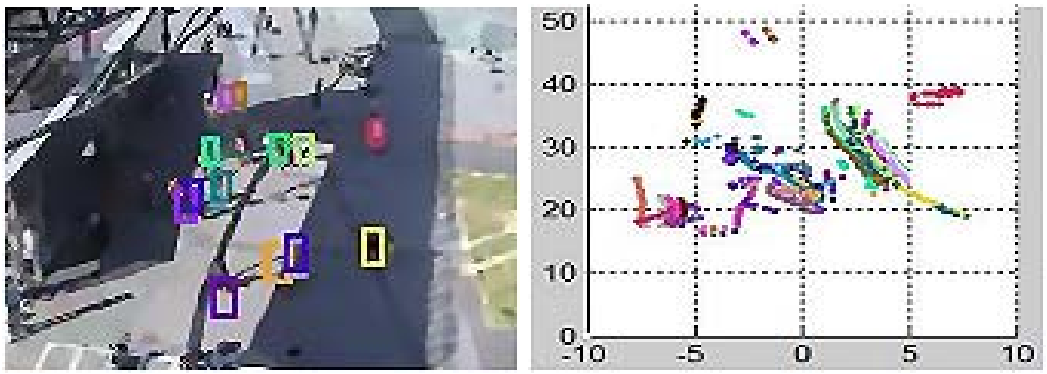
Figure 6.13: Deterministic tracking for the Monaco sequence. The color represents the tracker identity.



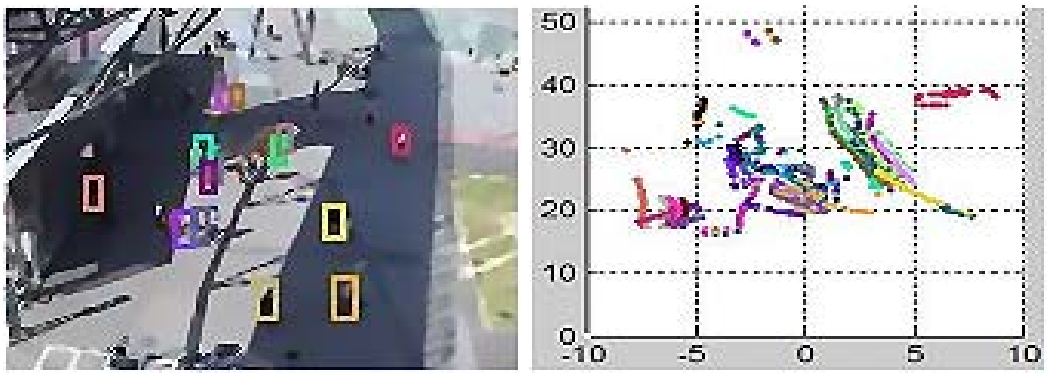
(a) flon 20



(c) flon 35



(e) flon 50



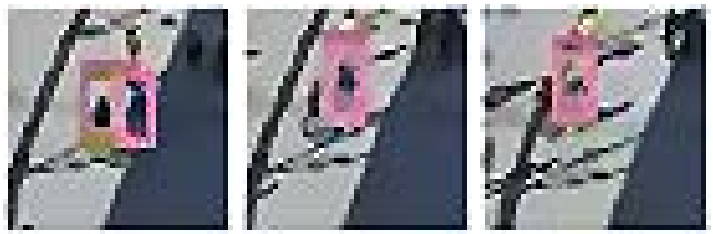
(g) flon 65

Figure 6.14: Deterministic tracking for the Flon sequence.

(a) no model 20 (b) no model 30 (c) no model 40 (d) no model 50

(e) model 20 (f) model 30 (g) model 40 (h) model 50

Figure 6.15: First example from the Flon sequence. Figures a,b,c,d refer to a pure correlation-based tracker. Figures e,f,g,h refer to the model-based tracker.



(a) no model 20 (b) no model 30 (c) no model 40

(d) model 20 (e) model 30 (f) model 40

Figure 6.16: Second example. The violet tracker without the model (on the left in figure a) jumps to the right losing one target.

Chapter 7

Automatic counting of pedestrians in video sequences

In Part I and Chapter 6 the main points of our work have been presented. A framework based on discrete choice analysis has been defined to model pedestrian walking behavior, looking at a trajectory as a sequence of choices over time. The information provided by the model consists in a discrete set of probabilities, describing the space around a current pedestrian position. A basic correlation-based tracker has been integrated with this model, showing that *behavioral prior information* is an interesting alternative to other dynamic models used for tracking, establishing a mathematical framework suitable to integrate higher level information, through a random utility and decision theoretic set up.

In this chapter, we deal with the output of the model-based tracker. The empirics and simplifications introduced to make the target detection step easier have raisen an interesting problem, which is actually present independently from the approach used to track targets. In Figure 1.2 we call the module in output of the tracking system as *target counting*. More precisely, assuming to have a tracking algorithm that provides an over-estimated number of targets, we investigate here some methods aiming to *reduce the bias* between the real number of targets which are present in the scene, and the target number as estimated by the tracking system. The problem is defined in Section 7.1 and a hierarchical clustering of pedestrian trajectories is proposed. In Section 7.2 we describe the different data representations, distance/similarity measures and grouping rules adopted to perform trajectory clustering. Finally, in Section 7.3 we report the obtained results.

7.1 Problem definition

From an ideal point of view, we would like to have a tracking system able to count the exact number of targets. The overestimation problem is related, but not identical, to the false positive problem for detection/tracking systems. Namely, a false positive is informally defined as a tracker placed on an image region, that does not correspond to a target of interest, which can be a background region or another object that we do not want to track. On the contrary, a target overestimation occurs when a target of interest is subject to a multiple detection, giving rise to multiple trackers, all of them being *correct*. In the context of pedestrian tracking, the overestimation of targets gives rise to the generation of multiple trajectories, related to the same individual. In Figure 7.1 we report an

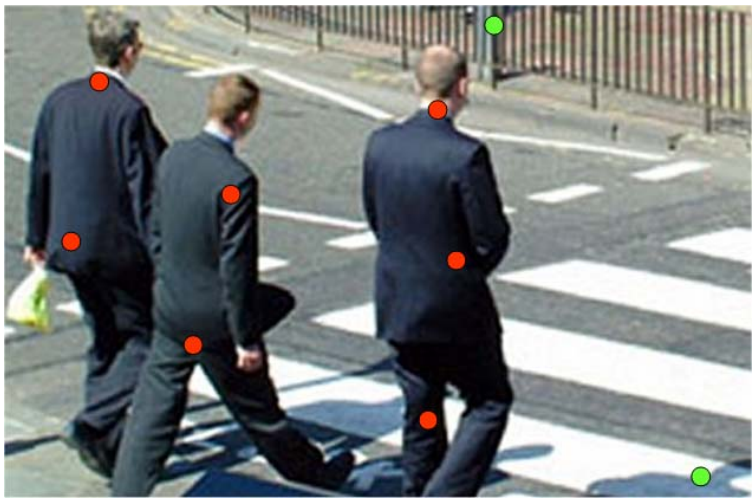
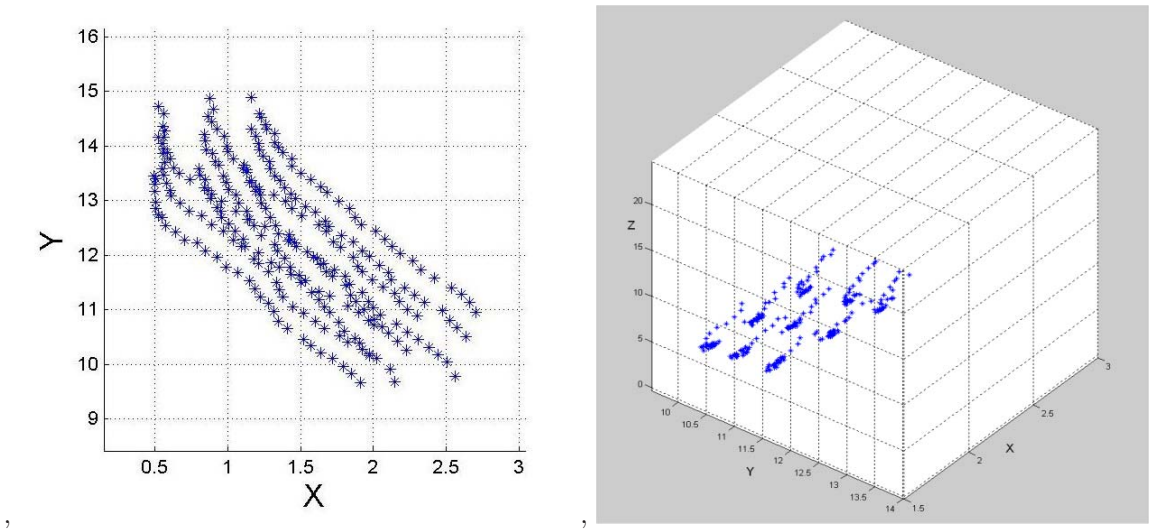


Figure 7.1: Overestimations vs false positives

example of overestimation (red markers) and false positives (green markers).
In Figure 7.2 we show a situation where multiple trackers are manually placed on 3 different individuals walking together, and manually tracked for a certain number of frames. Three trackers are place on the head, the center of the body and on the feet, respectively. Informally speaking, in Figure 7.2(a) it is hard to distinguish that the resulting 9 trajectories arise from 3 individuals. Figure 7.2(b) gives us a different viewpoint on the data, after adding the time dimension and having rotated the axes of the reference system.



(a) 2D representation

(b) 3D representation

Figure 7.2: An example of overestimated trajectories

We finally illustrate in Figure 7.3 how the same trajectory dataset looks like, after that a combination

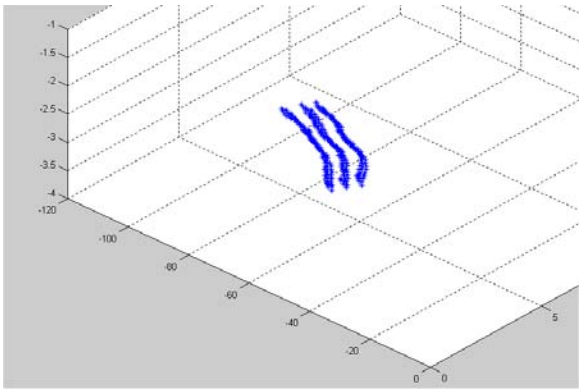


Figure 7.3: The same dataset after the application of a combination of linear transformations

of linear transformations (i.e., Independent Component Analysis (ICA) plus rotation) has been applied.
We note that now it is easier to recognize that the 9 trajectories belong to 3 well defined clusters, corresponding to the original 3 individuals. This reasoning allows to give a general formulation of the problem as an optimization problem.

Definition 7.1 *Given a trajectory dataset* $T = \{(x_i, y_i, t_i)\} \subseteq R^3$ *generated by a tracking system, with* $i = 1, ..., N$ *where* N *is the total number of observations, and given a clustering algorithm* r *between the trajectories and an objective function* J_r *measuring the performances of the clustering algorithm, we are interested in finding the mapping* $M : T \rightarrow T^{'} \subseteq R^3$ *maximizing* J_r.

The idea behind this general formulation of the problem is that counting targets from trajectories can be actually seen as providing a set of suitable (and general) transformations on the original dataset. Moreover, for a given association rule between the data, we are guaranteed to have the maximum discriminant power in the data association process. Of course, such a general formulation is intractable, and it represents more a qualitative description of an intuitive process than a mathematical definition. As a consequence, several simplifying assumptions have to be made, in order to make the problem operational.

7.1.1 Multi-layer hierarchical clustering approach

A natural way to scale down the complexity of the general optimization problem proposed above is to keep the main idea that those trajectories, originated from trackers who belong to the same target, are similar to each other. Moreover, the set of suitable transformations on the data is reduced to a specific and well-founded set of different data representations. This intuition leads to a reformulation of the problem in terms of a pure trajectory clustering problem. Finding the number of clusters in a trajectory dataset that overestimates the targets, would actually correspond to give an estimation of the number of individuals who generate the dataset itself. A similar idea, even if related to a different context, is used by Shechtman and Irani, 2005. The authors introduce a behavior-based similarity measure, telling whether two different space-time intensity patterns of two different video segments could have resulted from a similar underlying motion field. This is done directly from the intensity information, without explicitly computing the underlying motions. Moreover, this approach looks for correlated spatio-temporal patterns in a more general context, not specifically referred to pedestrians.

Specifically to the clustering domain, many research efforts have been made in the last three decades. A huge amount of literature exists on this subject and a lot of different methods have been defined. While a full review of the problem is clearly out of the scope of this thesis, we focus the attention on the main aspects of clustering, given just the "flavour" on this data analysis methodology. From a general point view, any clustering method is based on three main steps:

- data representation
- distance/similarity measures between patterns
- the choice of a grouping rule

Data representation

This first step in the general clustering problem can be identified with a general mapping (or a set of transformations) $M : T \rightarrow T^{'} \subseteq R^3$ applied to the original data set, in order to obtain a more discriminant data representation for cluster analysis. The transformations to apply in order to achieve this goal are strongly data dependent. We can say that normally a good data mapping has to provide a better partition of the samples as well as a dimensionality reduction, for computational efficiency. Just for the sake of clarity we give the following basic definitions:

Definition 7.2 *A pattern (feature vector, observation) X is a single data item used by the clustering algorithm. It consists of a vector of d measurements:*

$$X = (x_1, ..., x_d) \tag{7.1}$$

The individual scalar components x_i with $i = 1, ..., d$ are collected features (or attributes) and d is the dimensionality of the pattern space.

Definition 7.3 *A pattern-set S is a set*

$$S = (X_1, ..., X_n) \tag{7.2}$$

The i^{th} pattern in S will be denoted as $X_i = (x_{i,1}, ..., x_{i,d})$ and S is viewed as an $n \times d$ pattern matrix

All the feature selection/extraction methods belong to the set of transformations that map the original dataset into a more suitable representation. In the case of image data for example, edges, textures, color etc..., represent features that can help to improve the original representation. Dealing with time series, we could be interested in removing the offset or the trend from data, re-scale or again looking for recursiveness of patterns in time (Agrawal et al., 1995 and Box and Jenkins, 1970). Most of the time, a statistical generative model is given for the data (e.g., Principal Component Analysis (PCA) or Independent Component Analysis (ICA)) which can be used to change the coordinate system and achieve dimensionality reduction (Bell and Sejnowski, 1995, Stone, 2002 and Hyvärinen and Oja, 2000). If linear models do not well adapt to the data at hand, non-linear dimensionality reduction techniques such as Local Linear Embedding (Saul, 2000) and ISOMAP (Balasubramanian et al., 2002 and Souvenir and Pless, 2005) can be applied. There are no general guidelines suggesting methods to obtain a good data representation. The experience of the analyst and the data at hand represent the main sources of information.

Distance/similarity measures between patterns

Clustering approaches are based on fuzzy concepts, such as *nearness* or *relatedness*. To quantify these ideas, the choice of a distance and/or a similarity measure between patterns is necessary. The most popular metric for continuous features is the Euclidean distance, which is a special case of the more general Minkowski metric

$$d_p(x_i, x_j) = (\sum_{k=1}^{d} |x_{i,k} - x_{j,k}|^p)^{(1/p)} \tag{7.3}$$

when $p = 2$. This family of metrics work well when the data set has isolated clusters, but often they need a normalization of the data to avoid that the largest-scaled sample dominates the others. Another well known distance is the Mahalanobis metric, defined as

$$d_M(x_i, x_j) = (x_i - x_j)\Sigma^{-1}(x_i - x_j)^T \tag{7.4}$$

where x_i and x_j are assumed to be row vectors and Σ is the sample covariance matrix (or the known covariance matrix of the pattern generation process). d_M is optimal when the class conditional densities are multivariate Gaussian distributions. Recently, interesting approaches are those proposed in Eiter and Mannila, 1997 and Ramon and Bruynooghe, 2001 where similarity measures and metrics are defined based on the definition of specific relations between sets of points.
A classical approach, widely used in time-series analysis, is the DTW (Dynamic Time Warping, see Berbdt and Clifford, 1994; Keogh and Pazzani, 2000). The main idea behind DTW is to find an alignment of two time series on a common time-axis. Another classical approach is to use a vector-form for trajectories and use a p-norm to compute distances (see Yi and Faloutsos, 2000). This method does not tackle the problem of outliers, while most of the metrics used to compare data sets are sensitive to this phenomenon. A lot of work has been performed in the data mining community, mainly focusing on finding better distance measures to indexing items in databases (Agarwal et al., 2000 and Perng et al., 2000). Recently, several researchers have used the Hausdorff distance in a point set matching context (Eiter and Mannila, 1997 and Guo et al., 2003), while in the database retrieval domain an interesting similarity measure is the Longest Common SubSequence (LCSS) (Vlachos et al., 2002). We describe these two metrics more in details in the next chapter.

The grouping rule

A *class* is defined as a source of patterns, whose distribution in the feature space is governed by a probability density, specific to the class. Clustering techniques group patterns in such a way that classes thereby obtained reflect the different pattern generation process. In Figure 7.4 a summary of clustering techniques based on different grouping rules is illustrated.
Hard clustering approaches (King, 1967, Anderberg, 1973, Dubes and Jain, 1976, Jain and Dubes, 1988) assign a class label l_i to each pattern x_i, identifying its class. The set of the labels for a pattern set S is $L = (l_1, ..., l_n)$ with $l_i = 1, ..., k$ where k is the number of clusters. Fuzzy clustering procedures assign to each input pattern x_i a fractional degree of membership f_{ij} to each output cluster j (Zadeh, 1965, Bezdek, 1981, Dave, 1992). Hierarchical clustering approaches produce a nested series of partitions, based on a criterion for merging or splitting clusters. Such methods are more suitable in those cases where no *a priori* knowledge provides information on the number of clusters. The first family of these algorithms, *agglomerative*, begins with each pattern in a distinct (singleton) cluster, and successively merges clusters together, until a stopping criterion is satisfied (Strehl and Ghosh, 2000 and Wallace and Kollias, 2004). The second, *divisive*, begins with all patterns in a single cluster and performs splitting until a stopping criterion is riched (Boley, 1998).

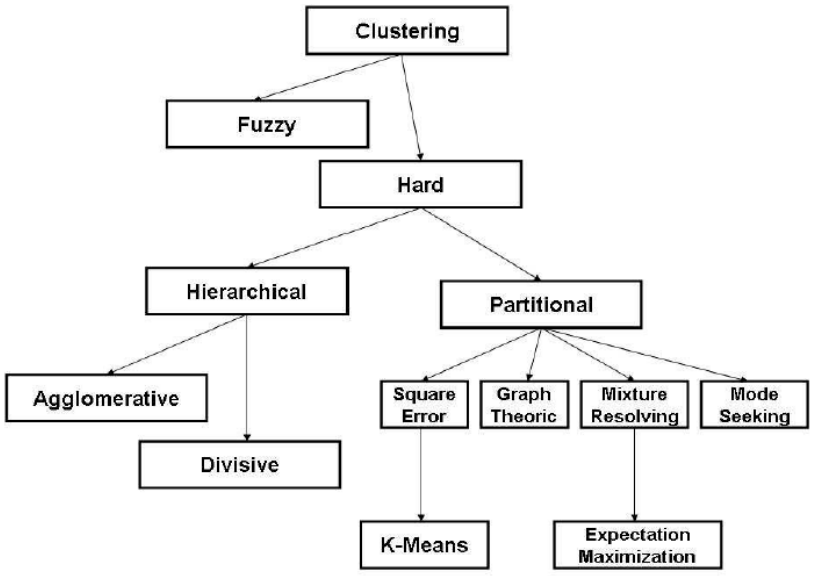


Figure 7.4: Overview on the clustering techniques, based on different grouping rules

Partitional clustering algorithms divide data in a certain number of groups, optimizing a clustering criterion (Mao and Jain, 1996; Symon, 1977; Dubes and Jain, 1976). The choice of the number of groups is made based on the *a priori* knowledge on the data at hand. Additional techniques for the grouping operation include probabilistic methods, where the underlying assumption is that the patterns to be clustered are drawn from one of several distributions. The goal is to identify the parameters of each of such distributions. Most of the work has been done assuming a maximum likelihood estimation for Mixture of Gaussians distributions (Mitchell, 1997; Jain and Dubes, 1988).

In the next section we present the application of different clustering techniques to pedestrian trajectories, in order to reduce the bias between the real number of individuals in the scene and the targets' number as estimated by our tracking system. Again, the problem of automatic target counting is independent from the algorithm used for tracking. Our aim is to provide a simple and well-founded post-processing method.

7.2 Clustering of trajectories for automatic counting of pedestrians

In this section different data representations and distance/similarity measures are presented, under a common hierarchical clustering framework. In Section 7.2.1 we introduce the general idea describing how the different techniques are combined, in Section 7.2.2 the different data representations are described, in Section 7.2.3 the different distance/similarity measures that have been tested are formally defined, and in Section 7.2.4 the grouping rule is described.

7.2.1 Proposed methods

A multi-layer structure

The idea for a multi-layer hierarchical clustering arises from the consideration that the comparison between trajectories can be performed from different point of views. Trajectories of different lengths rarely belong to the same person. Moreover, paths belonging to the same target likely start from close spatial points. In figure 7.5 we illustrate the conceptual tree structure of our clustering method.

First level Represents a length-based clustering, where trajectories having the similar length are grouped together. It is actually possible that pedestrians stay in the scene for different amount of time, yielding to trajectories of different lengths

Second level Individuals enter the scene at different spatial locations. A pre-clustering on the trajectory starting positions is performed, in order to separate different groups. We assume here that those trajectories belonging to the same pedestrian and/or to close individuals walking together start at close spatial positions.

Third level While the first two levels represent simple pre-processing operations on the original dataset, the actual counting task is performed at the third level. The discrimination between oversampled pedestrians and different individuals walking close to each other requires a more detailed analysis.

We approach the problem comparing and testing different data representations and distance/similarity measures, under a common hierarchical clustering framework.

Compared approaches

In Section 7.1.1 the clustering problem is identified with the choice of a data representation, a distance/similarity measure and a grouping rule. In Table 7.1 we report the different techniques that have been combined and tested.

Data representation	**Distance/similarity measure**	**Grouping rule**
ICA	Euclidean	Hierarchical agglomerative
TS	Longest Common SubSequence (LCSS)	
MCC	Hausdorff	
ICA = Independent Component Analysis TS = Time series MCC = Maximum of cross-correlation		

Table 7.1: The set of different data representations and distance/similarity measures that have been combined and tested, under a common hierarchical agglomerative clustering framework.

- *Clustering with ICA and time series representations* The aim here is to compare the two representations using both the Hausdorff distance and the LCSS similarity measures (Antonini and Thiran, 2004). ICA is a generative statistical model, indicated for clustering analysis on sparse data. It reduces the influence of outliers, grouping the data around the independent components. The goal is to show that a distance measure sensible to the presence of outliers

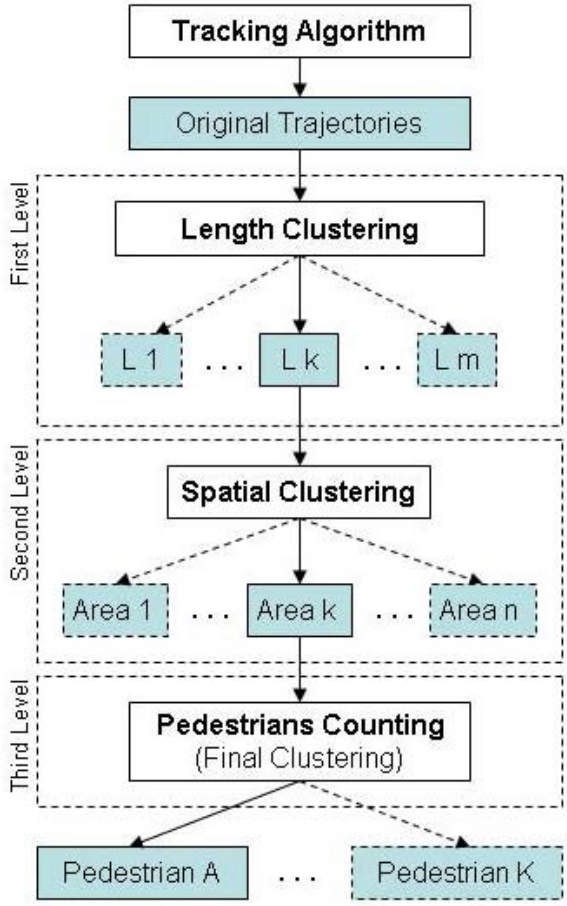


Figure 7.5: An overview of the proposed multi-layer clustering

(the Hausdorff distance) performs well, if used with a suitable representation. The time series representation does not reduce the presence of outliers, requiring a more complex similarity measure, such as the LCSS. The different combinations are tested on two different datasets, and the results are reported in Section 7.3.1.

- *Clustering with the MCC representation* A new representation is proposed, based on the cross-correlation between pairs of trajectories (Biliotti et al., 2005). The idea is that two identical trajectories are equally distant from a reference one. Mapping pairs of trajectories with their maximum correlation value allows to reduce the dimensionality of the data to a set of 3D points, where spatially close points represent trajectories which are similar to a reference one. We use the Euclidean distance with the MCC representation, testing the method on two datasets. The relative results are reported in Section 7.3.2

7.2.2 Data representations

Time series

A trajectory dataset in its original representation can be considered as a time series of 2D spatial points. Each point is represented by a *triplet* (x, y, t), the two plane coordinates (x, y) and the time step t. Time series analysis and clustering are important topics in many different context, from finance to bio-informatics. Several techniques exist, each of them presenting drawbacks and advantages when seen from different points of view. For our purposes, we discuss here two common pre-processing techniques for time series that we have use to pre-process the original data when using this representation. The first one is the *linear trend* removal. The trend in a time series represents the mean slope and can be computed with standard techniques, such as linear/non-linear regression. Intuitively speaking, removing the trend can be considered as a way to highlight fluctuations around the mean slope. The advantage of trend removal is that slight non-stationarities can be (partially) addressed. In the case of pedestrians walking in normal (no panic) conditions, we can expect a priori a certain degree of regularity and highly non linear time series should be unlikely (see Schadschneider, 2002; Helbing et al., 2002). As a consequence, we use a linear regression model to estimate and remove the linear trend. Another family of techniques widely used working with time series is represented by *smoothing* algorithms. These techniques are used to remove irregularities in the data and provide a clearer view of the underlying behavior of the series. When the trend has been removed a *single smoothing* algorithm can be used

$$f_t = \alpha y_{t-1} + (1 - \alpha) f_{t-1} \tag{7.5}$$

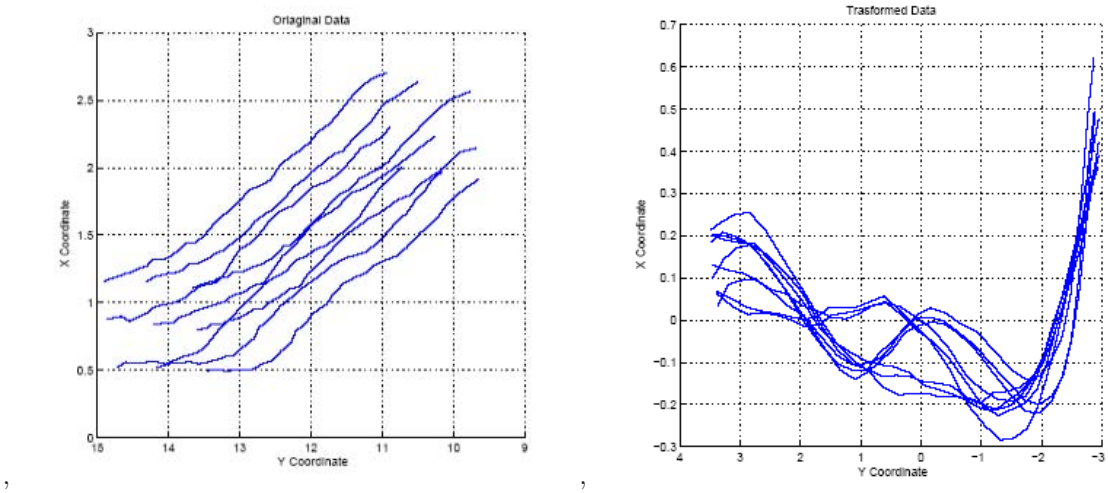
with $0 < \alpha \leq 1$ and $t \geq 3$. The α parameter is called the smoothing constant, f_t is the smoothed value and y_t the original value of the series at time t. We can also perform smoothing accounting for the trend at the same time, using *Double exponential smoothing*. The equations describing the model are

$$f_t = \alpha y_t + (1 - \alpha)(f_{t-1} + b_{t-1}) \tag{7.6}$$

$$b_t = \gamma(f_t - f_{t-1}) + (1 - \gamma) b_{t-1} \tag{7.7}$$

where the same notation as before has been used and b_t represents the trend at time t. The first smoothing equation adjusts f_t directly for the trend of the previous period, b_{t-1}. The second smoothing equation then updates the trend, which is expressed as the difference between the last two values. The equation is similar to the basic form of single smoothing, but here applied to the

updating of the trend. α and γ are the two smoothing constants, used to smooth the observation and the trend, respectively. They are are bounded in the interval $[0, 1]$. We report in Figure 7.6 an example of the effect of such a pre-processing techniques on the same manually tracked trajectory dataset shown in Figure 7.2(a).



(a) Original 2D data

(b) Pre-processed data

Figure 7.6: Nine trajectories are generated by 3 individuals. In the right-hand figure the effects of the pre-processing techniques. The trajectories are better grouped into three bundles

Independent Component Analysis

The main idea here is to consider trajectories as sequences of 3D points, (x, y, t), generated by a stochastic process. Walking pedestrians give rise to trajectories which are well different one from the other. Even if two persons follow the same spatial path, they do that at different times, leading the two trajectories to be separated when using a 3D representation. This fact leads the trajectory dataset to be sparse. These heuristics find a natural mathematical formalization in probabilistic generative models, which are well known in literature, widely used in almost any scientific domain involving statistical computation and analysis. Independent Component Analysis (ICA, Bell and Sejnowski, 1995; Stone, 2002) in particular is a generative model where a set of random variables, the *observations*, are supposed to be generated by a mixing process, starting from another set of statistical independent latent (unobservable) variables, the *sources*, by means of an unknown mixing matrix A. This model can be described by the following equation:

$$\mathbf{X} = A\mathbf{s} \tag{7.8}$$

where X represents the observations, s the sources and A is the mixing matrix. The number m of observations can differ from the number n of sources. For a general discussion on ICA we can assume, without loss of generality, that $m = n$. The basic hypothesis of the ICA model is the statistical independence of the latent variables. This property can be derived using an information-theoretic framework. We define the mutual information I between m scalar random variables $y_i, i = 1, ..., m$ as follows:

$$I(y_1, ..., y_m) = \sum_{i=1}^{m} H(y_i) - H(\mathbf{y}) \tag{7.9}$$

where H represents the differential entropy. The mutual information is equivalent to the Kullback-Leibler divergence between the joint density of $\mathbf{y}$ and the product of the marginal densities of the y_i. This measure is zero if and only if the variables y_i are statistically independent. It is possible to show that constraining the y_i to be uncorrelated and of unit variance, the mutual information is equal to:

$$I(y_1, ..., y_m) = C - \sum_{i} J(y_i) \tag{7.10}$$

where J represents the negentropy, defined as:

$$J(\mathbf{y}) = H(\mathbf{y}_{gauss}) - H(\mathbf{y}) \tag{7.11}$$

It is well known from information theory that Gaussian variables have the maximum entropy among all the variables with equal variance. We obtain that minimizing the mutual information is equivalent to maximize the negentropy, which actually means to maximize the *non Gaussianity* of the random variables (see Hyvärinen and Oja, 2000). So, the main assumption in ICA is the non-Gaussianity of the source signals.

Geometrical interpretation ICA becomes interesting for our purposes when we consider its geometrical interpretation, compared to Principal Component Analysis (PCA). While the PCA solution is given by orthogonal axes representing the directions of maximum variance in the data, ICA can be seen as the non-orthogonal extension of PCA. In Figure 7.7 this property is illustrated (Figure 7.7 is taken from Bartlett et al. (2002)).

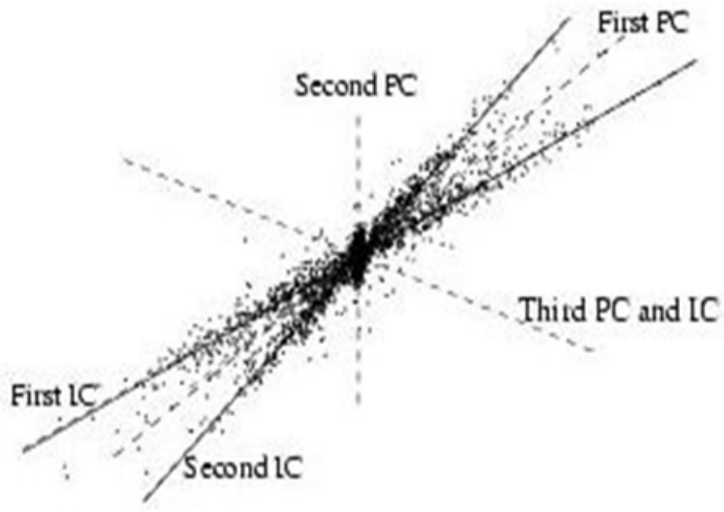


Figure 7.7: ICA vs PCA

When the sources are sparse, ICA provides a better probabilistic model of the data than PCA, which better identifies where the data concentrate. The chosen solution is based on the high-order

statistics of the data and represents a non-orthogonal rotation. As a consequence, this transformation can change the relative distances between points affecting similarity and/or distance measures. For these reasons, it can be quite useful in classification and clustering problems. Figure 7.2(b) and Figure 7.3 illustrate an example of trajectories projected in the ICA space.

The limitation of this representation resides in an ambiguity intrinsic in the ICA model. In equation 7.8, both $\mathbf{s}$ and A are unknown. We can change the order of the independent components keeping untouched the validity of the model. Therefore, the components are estimated up to a permutation matrix. When the ICA model is used, for example, as a dimensionality reduction method (by means of a previous PCA step, where a certain number of eigenvalues of the covariance matrix are kept) this doesn't change the results. On the contrary, in our case we use the ICA model to estimate a transformation matrix, changing the space where the data are represented. Permuting the order of the estimated components is the same as inverting the axis of the new representation system, changing the data representation itself. This fact leads to different clustering results. One solution can be to keep the ICA estimation that optimizes the clustering. In our specific case, having three independent components, the number of permutations is 3!. As a consequence, it is possible to choose the order which maximizes the clustering performances. This ambiguity in the ICA model can seriously deteriorate the performances when such a model is applied to high dimensional datasets, where the number of permutations become huge.

Maximum of cross-correlation

We introduce here the maximum of cross correlation (MCC) representation. The idea is simply the realization that two identical trajectories are always equally far from a reference one. This simple fact is used here. We fix any trajectory t_1 of the dataset as the reference trajectory. We compute the similarity measure between two trajectories as the cross-correlation function between them. We can look at two trajectories t_1 of length M and t_2 of length N as two real 2D discrete signals, and write the cross-correlation function c between them as:

$$\begin{aligned} &c(m,n) = t_1(-m,-n) * t_2(m,n) = \\ &\sum_{j=0}^{M-1}\sum_{k=0}^{N-1} t_1(-j,-k)t_2(m-j,n-k) \end{aligned} \tag{7.12}$$

The two trajectories are represented by two matrices of size Mx2 and Nx2 respectively, so the size of the full cross-correlation is $(M+N-1)$x3. We show in figure Figure 7.9(a) the 3D representation of the output c where the axes represent the three columns of the cross-correlation. The new trajectory representation is obtained mapping each pair of trajectories with the *maximum* of their cross-correlation. The intuitive idea is that, independently from the chosen reference trajectory t_1, the maximum of the cross-correlation between two *similar* trajectories t_2 and t_3 with t_1 maps t_2 and t_3 into two close spatial points. In a similar way, two strongly different trajectories will be mapped into two farther spatial points. In Figure 7.8 a new set of 30 trajectories, manually tracked from 10 pedestrians, is illustrated. The individuals are walking in 3 different groups composed by 3, 3 and 4 persons, respectively. Looking at Figure 7.8 is easy to identify the 3 groups, but it is not easy at all to count the 10 pedestrians. Figure 7.9 and Figure 7.10 illustrate the 3D MCC with all the 2D projections.
This representation presents several advantages over the others. First, it can handle trajectories of different lengths, in a quite easy manner, being the cross-correlation operator independent on

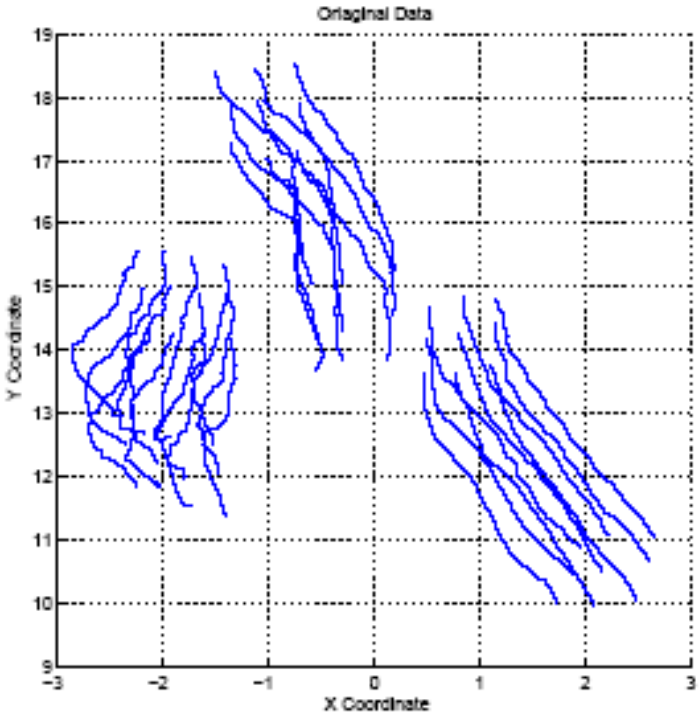
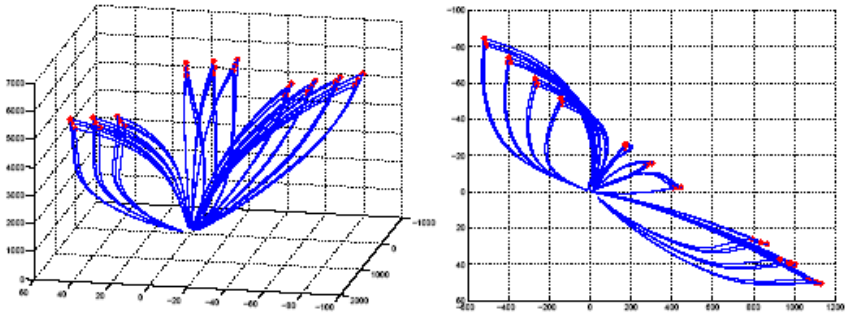


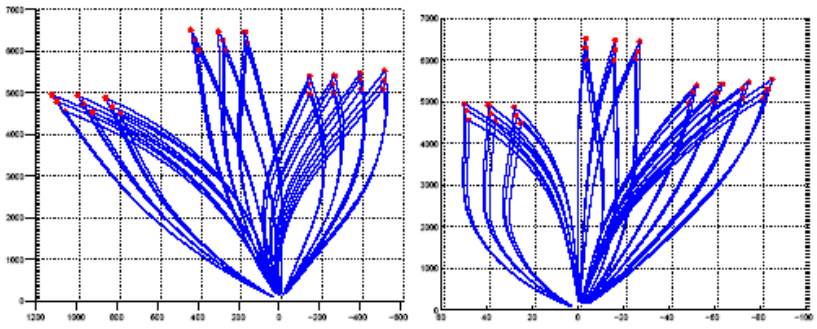
Figure 7.8: New set of 30 trajectories, manually tracked, corresponding to 10 pedestrians



(a) 3D view (b) x-y view

Figure 7.9: MCC in 3D and the x-y 2D projection

the number of points. Second, it allows to map a couple of trajectories into one 3D point. This remains true for any dimensionality of the dataset and represents a drastic dimensionality reduction. Third, it allows to reduce the clustering problem to a much simpler spatial clustering, which can be handled, with a certain accuracy, by means of the simple Euclidean metric.



(c) z-x view (d) z-y view

Figure 7.10: The other 2D projections

7.2.3 Distance/similarity measures

Hausdorff distance

The Hausdorff distance is a metric between non-empty compact point sets. Let $X_1 = (x_{11}, ..., x_{1m})$ and $X_2 = (x_{21}, ..., x_{2n})$ be two finite point sets. The Hausdorff distance $H(X_1, X_2)$ is defined as follows:

$$H(X_1, X_2) = \max(h(X_1, X_2), h(X_2, X_1)) \tag{7.13}$$

where $h(X_1, X_2)$ is the *direct* Hausdorff distance between X_1 and X_2, defined as

$$h(X_1, X_2) = \max_{x_1 \in X_1} D(X_1, X_2) \tag{7.14}$$

where $\forall x_1 \in X_1$, $D(x_1, X_2)$ is defined as

$$D(x_1, X_2) = \min_{x_2 \in X_2} d(x_1, X_2) \tag{7.15}$$

It identifies the point $x^* \in X_1$ that is farthest (using a pre-specified norm d, usually the Euclidean distance) from any point in X_2 and measures the distance from x^* to its nearest neighbour in X_2. Essentially, $h(X_1, X_2)$ ranks each point in X_1 based on its distance from the nearest point in X_2 and then uses the largest ranked such point (x^*) as the distance measure. Similarly, we can define $h(X_2, X_1)$. The Hausdorff distance is the maximum between the direct and inverse distances. As it is well known, this metric is very sensitive to outliers so smoothing operations or other kind of transformations, as for example the ICA representation, are usually performed before to compute the distance. On the other hand it has also some quite good properties. First, it represents a metric and not just a similarity. Second, we can easily apply this measure to sets of different sizes.

Longest Common SubSequence

Longest common subsequence (LCSS) is a similarity measure derived from the Levenshtein distance, also known as edit distance measure (Levenshtein, 1966). The edit distance is a measure of the similarity between two strings, given by the number of deletions, insertions, or substitutions required to transform one string into the other. In this spirit, and using the notation used in Vlachos et al. (2002), we use what the authors call the $S1$ similarity measure. It does not extend to translations

because in our case two parallel trajectories with similar shapes may represent two different individuals.
Given two trajectories $A = ((a_{x,1}, a_{y,1}), ..., (a_{x,n}, a_{y,n}))$ and $B = ((b_{x,1}, b_{y,1}), ..., (b_{x,m}, b_{y,m}))$, let $Head(\mathrm{A})$ and $Head(\mathrm{B})$ be two sequences defined as:

$Head(A) = ((a_{x,1}, a_{y,1}), ..., (a_{x,n-1}, a_{y,n-1}))$
$Head(B) = ((b_{x,1}, b_{y,1}), ..., (b_{x,m-1}, b_{y,m-1}))$.

Definition 7.4 *Given an integer $\delta \geq 0$ and a real number $0 < \epsilon < 1$ the $LCSS_{\delta,\epsilon}(A, B)$ is defined as follows:*

$$\begin{cases} 0 \text{ if } A \text{ or } B \text{ is empty} \\ \\ 1 + LCSS_{\delta,\epsilon}(Head(A), Head(B)), \\ \text{if } |a_{x,n} - b_{x,m}| < \epsilon \text{ and } |a_{y,n} - b_{y,m}| < \epsilon \text{ and } |n - m| \leq \delta \\ \\ \max(LCSS_{\delta,\epsilon}(Head(A), B), LCSS_{\delta,\epsilon}(A, Head(B))), \\ \text{otherwise} \end{cases}$$

Definition 7.5 *Given two trajectories A and B and given $\epsilon \in (0, 1)$ and $\delta \geq 0$, the similarity measure $S1$ is defined as follows:*

$$S1(\delta, \epsilon, A, B) = \frac{LCSS_{\delta,\epsilon}(A, B)}{\min(m, n)} \tag{7.16}$$

The constant δ controls how far in time we can go in order to match a given point from one series to a point in the other time series. ϵ is the matching threshold. LCSS similarity has the very nice property of matching two sequences stretching them, without rearranging the order and allowing for some *unmatched* elements. This is not allowed for example using Euclidean distance or DTW, which require all the elements to be matched, including the outliers. For this reasons, LCSS is normally better in presence of outliers.

7.2.4 Grouping rule

Our aim is to reduce the bias in the number of targets as estimated by the tracking system. We do not know a priori how many pedestrians are present in the scene. As a consequence, the hierarchical approach represents a natural way of grouping data over a variety of scales. We use both the *agglomerative* and *divisive* techniques.

- **Agglomerative:** trajectories are paired into binary clusters, the newly formed clusters are grouped into larger clusters until a hierarchic tree is created. The resulting tree can be analyzed at different scales, to find out different resulting data partitions. An agglomerative algorithm yields a *dendogram* representing the nested groups of trajectories and the similarity levels at which the grouping changes. Given n trajectories, the pairwise distance information is represented by a vector of length $n(n - 1/2)$. The linking method we use to generate the hierarchical tree is based on the average distance measures. Let u and v two clusters of size n_u and n_v respectively and let be x_{ui} the ith object in cluster u. We have:

$$d(u,v) = \frac{1}{n_u \cdot n_v} \sum_{i=1}^{n_u} \sum_{j=1}^{n_v} dist(x_{ui}, x_{vj}) \tag{7.17}$$

where the averaged pair distance between all the object pairs in the two clusters is used.

- **Divisive:** Hierarchical divisive clustering starts with a single cluster containing all the given objects and it keeps splitting the clusters based on some criterion in order to obtain a partition of singleton clusters. We report in the following the main steps of the used algorithm (Clason, 1990):

  1. from the whole set of trajectories we choose any one to be the first hub;
  2. find the trajectory which is farthest from this hub and make it the second hub;
  3. for each remaining trajectory, assign it to the closer hub;
  4. to decide for another hub:
     - find the average distance between the two hubs d;
     - compute the distance from each trajectory to its hub. If any distance is greater than d, define another hub for such a trajectory;
  5. the new hub is the trajectory which is farthest from its respective hub;
  6. re-compute the distance of each trajectory to the new hub and reassign the trajectory to the new hub in the case that the new distance is less then the distance with the previous assigned hub;
  7. repeat iteratively from step 4 until all trajectories are within a distance d from their hub or all points are themselves hubs.

In the previous chapter a general formulation for the counting pedestrian problem has been given. The optimization problem is intractable and simplifying hypothesis have to be done. In this chapter we have proposed some existent methods for data representation (time series and ICA) and we have proposed a new one (MCC) based on the simple mapping of two trajectories into a 3D point, represented by the maximum of the cross-correlation between the trajectories. Different distance/similarity measures have been chosen, depending on the method chosen for data representation. In the following chapter we present the experimental results obtained by the combination of the techniques defined here.

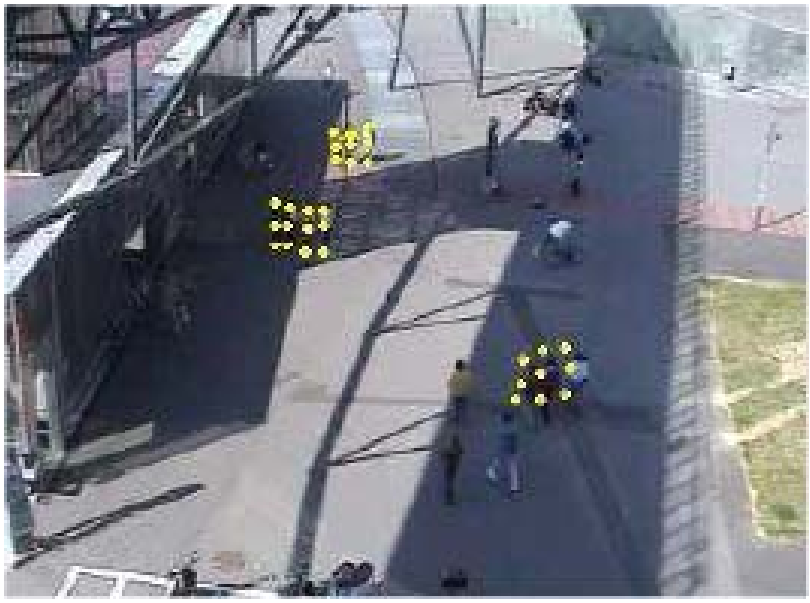
7.3 Results

We report in this section the quantitative results obtained applying the different trajectory clustering procedures to different datasets. All the trajectories coming from the tracker are considered in the clustering step, including those that not correspond to real individuals. Such an error comes actually from the tracking system and cannot be corrected with the proposed clustering approach. The first experiment, reported in Section 7.3.1, compares the clustering results obtained with time series and ICA representations, using the Hausdorff distance and the LCSS similarity measures (Antonini and Thiran, 2004). The second experiment in Section 7.3.2 compares the ICA representation with the MCC representation (Biliotti et al., 2005).

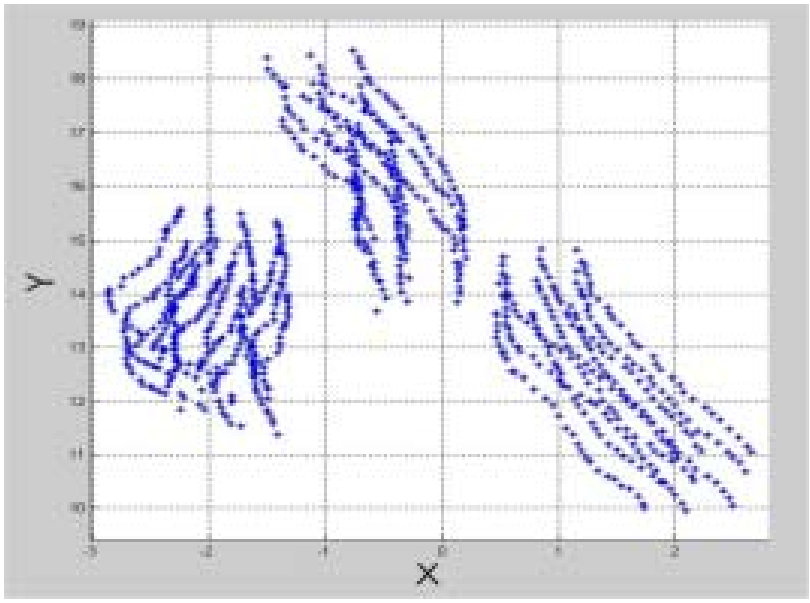
The results are compared defining two kind of errors. We call **e1** the number of *false negatives*, meaning that no clusters refer to an individual. We call **e2** the number of *false positives*, meaning those pedestrians having more than one resulting cluster over themselves.

7.3.1 Test 1

In this first test we compare the time series representation with the ICA representation. The Hausdorff distance and the LCSS similarities are used with both the representations. The aim of this experiment is to show that better results can be obtained using a more suitable data representation, which reduces the presence of outliers. Actually, also a metric as the Hausdorff one, extremely sensible to outliers, can perform well when it is used with an opportune data representation. Two sets of trajectories are used. The first one is composed by 30 trajectories manually grabbed and the second one consists in 15 trajectories obtained with our model-based tracking system. We show in the following the obtained results. The manually tracked points that generate our first data set are placed on 10 different pedestrians, 3 for each of them, and are placed on the head, the body's center and on the middle of feet of the individuals. The selected 10 pedestrians walk divided in groups of respectively 3, 3 and 4 persons, as we can see in figure Figure 7.11(a). The goal is to correctly cluster the 30 trajectories in 10 different groups. We show in figure Figure 7.11(b) the trajectories. The results on the first dataset are summarized in Table 7.2 and Table 7.3.



(a) Trackers used to collect the data



(b) Trajectories generated by the trackers

Figure 7.11: The first dataset used in Test 1

num traj	clustering alg	num clusters	num ped	e1	e2
30	*Time series with Hausdorff distance*	13	10	1	4
30	*Time series with LCSS similarity*	10	10	1	1

Table 7.2: Results obtained using the Hausdorff metric and *LCSS* similarity with a time series representation

num traj	clustering alg	num clusters	num ped	e1	e2
30	*ICA with Hausdorff distance*	10	10	/	/
30	*ICA with LCSS similarity*	10	10	/	/

Table 7.3: Results obtained using the Hausdorff metric and *LCSS* similarity in the ICA space

In Table 7.4 and Table 7.5 we report the results obtained using the second dataset.

num traj	clustering alg	num clusters	num ped	e1	e2
15	*Time series with Hausdorff distance*	6	6	2	2
15	*Time series with LCSS similarity*	1	6	5	/

Table 7.4: Results obtained using the Hausdorff metric and *LCSS* similarity with a time series representation

num traj	clustering alg	num clusters	num ped	e1	e2
15	*ICA with Hausdorff distance*	6	6	1	1
15	*ICA with LCSS similarity*	6	6	1	1

Table 7.5: Results obtained using the Hausdorff metric and *LCSS* similarity in ICA space

Table 7.2, Table 7.3, Table 7.4 and Table 7.5 present different interesting points to discuss. The results for the first data set clearly show how the ICA transformation improves the clustering. We can see it also in the respective results using the Hausdorff and LCSS metric/similarity. The differences of the respective results in the original space are removed in the ICA space, where the Hausdorff

distance performs as well as the LCSS similarity measure. This is an implicit indication that the non-orthogonal rotation has reduced the presence of outliers in the trajectories, concentrating the data along the independent directions. We remark the same qualitative improvements for the second data set.

7.3.2 Test 2

The results illustrated in the previous section show that a suitable data representation can overcome the drawbacks related to a specific metric. As we have already said in Section 7.2.2, the ICA representation presents some limitations, due to the nature of the ICA model itself. The independent components are defined up to a permutation matrix. This fact can create problems when we use such components to change the representation of our data. In this section we compare the ICA representation with the MCC. The first dataset used in this experiment consist in 31 trajectories distributed on 11 pedestrians (Figure 7.12(a)). The density of the targets in the scene is high. In particular, we note that the group of four pedestrians walking together (Figure 7.13(a)) is highly over-estimated by the detection/tracking algorithm. The numerical results are presented in Table 7.6. The clustering results on the trajectories are shown in Figure 7.12(b) and Figure 7.12(c) while visual examples are shown in Figure 7.13(b) and Figure 7.13(c).

num traj	**num clusters**	**num ped**	**e1**	**e2**
Independent Component Analisys:				
31	14	11	0	3
Cross-correlation:				
31	12	11	0	1

Table 7.6: Results for the *Flon* sequence.

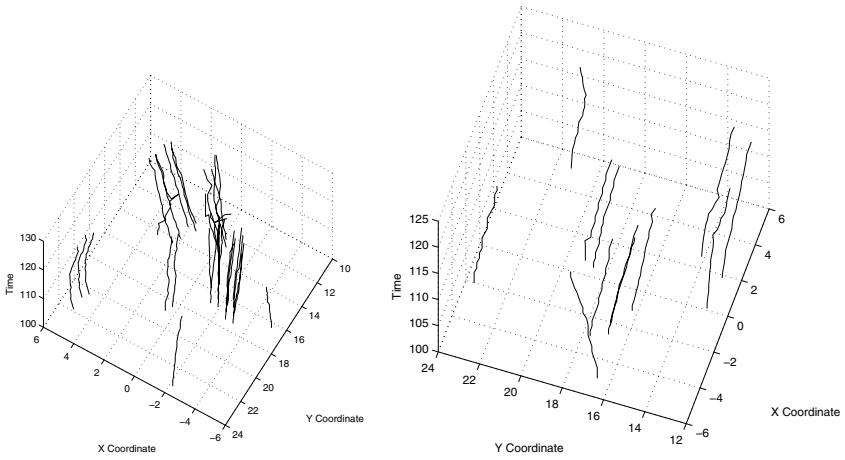
The second dataset used in this experiment is strongly over-estimated by the detection/tracking system. Eight pedestrians are present in the scene but the trajectories obtained are 43. We report in Table 7.7 the relative numerical results. The clustering results on the trajectories are shown in 7.14(b) and 7.14(c) while visual examples are shown in 7.15(b) and 7.15(c).

num traj	**num clusters**	**num ped**	**e1**	**e2**
Independent Component Analisys:				
43	17	8	0	4
Cross-correlation:				
43	9	8	0	1

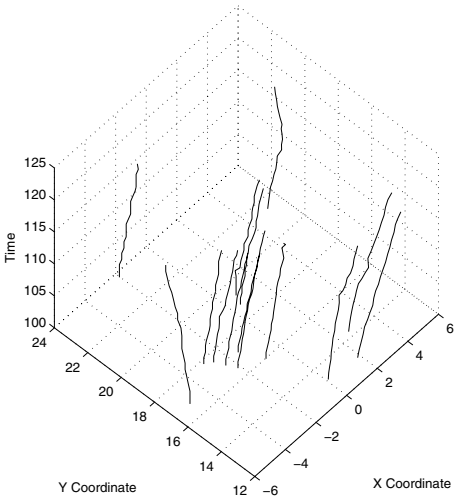
Table 7.7: Results for the *Monaco* sequence.

7.4 Summary

In this chapter we have presented a comparative study of clustering methods for automatic counting of pedestrians in video sequences. The aim is to reduce the bias in the real number of targets



(a) The *Flon* trajectory set (b) Cross-correlation-based clustering



(c) ICA-based clustering

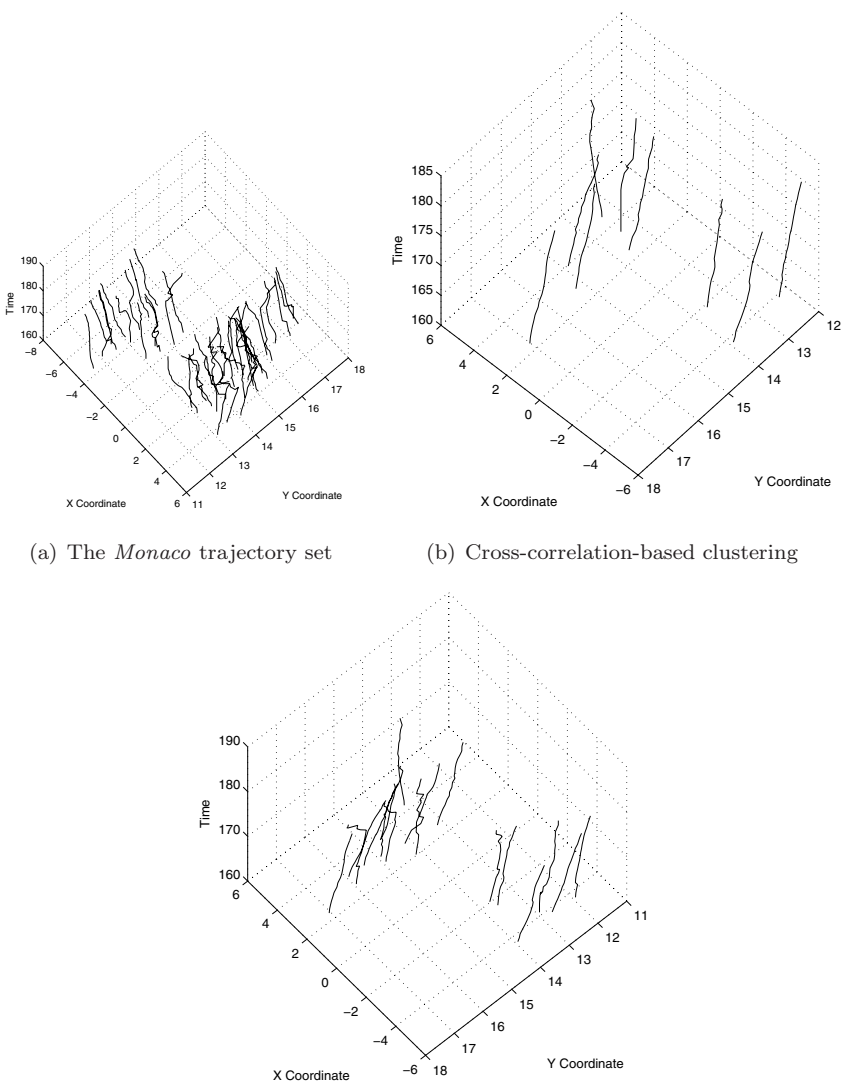
Figure 7.12: The results of the clustering on the *Flon* trajectory data set.

(a) The final trajectory points without clustering

(b) The final trajectory points after the max-of-cross-correlation clustering

(c) The same example after the ICA clustering

Figure 7.13: Visual examples for the *Flon* sequence.



(a) The *Monaco* trajectory set

(b) Cross-correlation-based clustering

(c) ICA-based clustering

Figure 7.14: The results of the clustering on the *Monaco* trajectory data set.

(a) The final trajectory points without clustering

(b) The final trajectory points after the max-of-cross-correlation clustering

(c) The same example after the ICA clustering

present in the scene, as estimated by the model-based tracker. Such system actually provides an over-estimation of the real number of targets. We do not focus on the errors coming from the detection/tracking steps but rather we attempt to exploit the information provided by it. At first, the data set is analysed based on the length and starting point position of trajectories. On the resulting *pre-clustered* dataset, different data representations and distance/similarity measures have been used. More specifically, we first apply both the Hausdorff distance and LCSS similarity for the time series and ICA representations. The results presented in Section 7.3.1 show that the ICA space provides a more suitable representation with respect to the original space-time domain, reducing the presence of outliers. The second experiment presented in Section 7.3.2 shows that the *maximum-of-cross-correlation* mapping allows for better clustering results providing at the same time dimensionality reduction at a low computational cost. The trajectory clustering problem is so reduced to a simpler 3D spatial clustering using the Euclidean metric. We finally comment on the fact that the clustering approach to count targets is independent from the algorithm used for tracking. The only condition is the target overestimation. The conclusions presented here indicate an alternative approach to detection and tracking. Computational resources can be saved in the detection step, allowing for multiple detection of targets. The resulting bias in the number of targets can be reduced in a post-processing step, by means of trajectory clustering techniques.

Chapter 8

Conclusions and future works

8.1 Research Summary

8.1.1 Discrete choice models for pedestrian walking behavior

The principal aim of this thesis was to investigate econometric tools, namely DCMs, for pedestrian walking behavior. The attention paid to these methods has been motivated by two main reasons. On one side, reliable results obtained in microeconomics and transportation science have shown how DCMs represent interesting descriptive models, designed to capture both correlation over alternatives and heterogeneity over the population of decision makers. In those cases of choices over a discrete set of alternatives, the random utility representation provides a consistent way to combine individual preferences with the knowledge of the modeller about the choice process. A rational behavior paradigm is assumed and it is implemented through a random utility maximization approach. The second reason to motivate our study is the always higher interest in model-based methods to pedestrian tracking. Scene analysis and activity recognition tasks are hard to perform in real scenarios using only standard image-based approaches. In this context, our aim was to provide new ideas and methodologies, integrating behavioral models for pedestrian dynamics with state-of-the-art detection/tracking techniques.

In Part I a general framework for pedestrian walking behavior based on DCM has been defined. The walking process is seen as a sequence of spatial locations choices over time. An *adaptive* model for spatial discretization has been given. The choice set is defined in a subjective way, depending on the current speed and direction of individuals. It represents a first way to take into account differences between decision makers. Two classes of behavioral patterns are identified: *constrained* and *unconstrained*. Constrained patterns describe the interactions with other individuals and they are further classified into *attractive* and *repulsive* interactions. The former are captured through a *leader follower* model. People are influenced by other individuals that have similar kinematic characteristics, when they are in a neighbour of the decision maker. When these individuals, i.e. leaders, accelerate or decelerate, a corresponding acceleration or deceleration as well as a direction change are induced on the decision maker. Its value depends on the kinematic characteristics of the leader, namely position, moving direction and speed. This phenomenon is important and it contributes to explain the formation of lanes in highly crowded conditions. The repulsive interactions are captured through a *collision avoidance* model. Similarly to the leader case, those individuals identified as col-

liders influence the walking behavior of the decision maker. Based on their kinematic parameters, the current pedestrian adapts her direction and speed, in other to avoid possible collisions. The unconstrained patterns capture the behavior of individuals, independently from the other people present in the scene. Pedestrians perceive as costs both unconstrained accelerations and changes in directions. In free-flow conditions, the acceleration behavior is modeled as a continuous function of the current speed. Elasticity parameters are introduced, showing that the attractiveness for both positive and negative accelerations are non-linear functions of the current speed value. A link is established with higher level decision making processes (tactical and strategical decisions). In the short-range context, the individuals move directly toward the final destination they have chosen at the strategical level. This behavior is captured through the *forward destination* model. A certain regularity of the walking trajectories is captured through the *keep direction* model, showing that individuals have the tendency to keep their current direction. These results are important in a pedestrian modeling context, being estimated on real data and validated by means of a pedestrian simulator.

8.1.2 A model-based pedestrian tracking system

In Part II the behavioral model is applied in a computer vision application, namely detection, tracking and counting of pedestrians in video sequences. The aim is to integrate the *prior* information provided by the model with the image-based information. Two methods are investigated, the first deterministic and the second probabilistic, under a common initialization scheme. Under the hypothesis of working with a monocular calibrated camera, the integration is done on the top-view plane. The target detection step is performed initializing the image through a sub-sampled foreground mask, obtained by background subtraction. Each hypothetical target is tracked by correlation, over a certain number of frames (evaluation period), and the resulting hypothetical trajectories are filtered using the calibrated model. The main advantage of this method is the simplicity. Well known problems related to the target detection step are bypassed, leaving to the model the role of selecting the most *human-like* trajectories. The result of this operation is a *dynamic detection* step, meaning that pedestrians are detected taking into account their behavior over the evaluation period. In the deterministic approach, pedestrians are tracked repeating the dynamic detection process. In the probabilistic approach, once the targets are detected, they are tracked over the video sequence using a Bayesian approach. The prior distribution is represented by the choice probabilities given by the model, while the likelihood term is equal to the normalized correlation. Tracking is performed by a *maximum a posteriori* approach.

8.1.3 Counting pedestrians

The simple initialization scheme proposed above gives rise to an overestimation of the number of targets. It is actually possible that more points of the sub-sampled foreground belong to the same target, generating multiple trajectories referring to the same individual. In this case, the filtering performed by the model does not discriminate between these trajectories, considering all of them as *human* trajectories. This problem has been tackled by a clustering approach. We assume that those trajectories belonging to the same target are more similar than trajectories belonging to different targets. The aim is to determine clusters of trajectories, where each cluster contains those trajectories generated by the same target. Several data representations are proposed and different distance/similarity measures are tested, under a common hierarchical clustering framework. In the first representation, trajectories are considered as time series of two-dimensional spatial positions.

In the second one, they are treated as three-dimensional points originated by a generative statistical process (ICA). Two different metrics, namely the Hausdorff distance and the Longest Common SubSequence similarity measure are tested on both the representations. Experimental results show that a metric notably sensible to outliers as the Hausdorff one, performs well when used with a more robust data representation, such as ICA. A third representation maps pairs of trajectories with a three-dimensional point, given by the maximum of their cross-correlation. This approach is simple, computationally convenient and it allows for a strong dimensionality reduction. Clustering on this representation is performed using the Euclidean distance, and experimental results show that it outperforms the other tested approaches.

8.2 Contributions

The proposed research is characterized by the cross-disciplinarity between behavioral modeling and computer vision. Three specific problems are treated: pedestrian walking behavior modeling, detection and tracking of pedestrians and counting pedestrians in video sequences. This thesis contributes to the state of the art of both these topics in the following respects:

- A new microscopic model for short-range pedestrian walking behavior, based on discrete choice analysis, is proposed. The framework provides a classification of the behavioral patterns into constrained and unconstrained behaviors. The constrained patterns are further classified as attractive and repulsive interactions. A dynamic and individual-based spatial representation, adaptive with individual speed and direction, is provided.

- Leader follower and collision avoidance models are defined to model the constrained behaviors. We assume that the decision maker identifies leaders and colliders in the space around, based on the similarities among the respective kinematic characteristics, namely speed, moving direction and position. These models are specified using a sensitivity-stimulus framework, widely used in transportation science (driver behavior modeling) for car following models. It has been adapted and extended to the pedestrian case, for both the leader follower and collision avoidance models.

- The unconstrained portion of the model takes into account the behavior of individuals, independently of the presence of other pedestrians. It captures the trade-off between three behaviors: the attractiveness for accelerations depends continuously on the current speed value; the tendency of individuals to keep their current directions; the tendency of individuals to move towards their final destination, chosen at a higher level in the decision process.

- The spatial correlation among the alternatives of the choice set is captured defining different error structures, based on a cross nested logit and a mixed nested logit specifications.

- The parameters of all components of the pedestrian walking behavior are estimated jointly using two different sources of pedestrian trajectory data, collected from video sequences recorded in real scenarios. The result is a model applicable to walking pedestrians in real, normal (no panic) conditions, rather than only specific cases. The heterogeneity in the data has been captured, using different error terms for the two sources of data, and estimating the related scale parameter.

- A detection/tracking system for pedestrians in video sequences, based on the proposed behavioral model, has been developed. In the proposed approach, detection and tracking are

interpreted as inter-operating steps. The idea of using behavioral prior information is implemented by the dynamic detection algorithm.

- Two approaches to track pedestrians over time are proposed. A deterministic one, where tracking is interpreted as a consecutive detection task and implemented repeating the dynamic detection step. A probabilistic one, where the behavioral and image-based information are combined under a Bayesian framework.

- An algorithm to count pedestrians in video sequences is proposed. It is based on hierarchical clustering of trajectories produced by the tracking system. It can be used independently from the tracking approach, under the assumption that the latter provides an overestimation of the number of targets.

8.3 Directions for future research

Several tracks for future research can be identified from this work.

- Pedestrian behavior is complex and a suitable framework to represent the decision making processes involved is the hierarchical one, proposed by Hoogendoorn (2003) and Daamen (2004). We have shown in this work that DCMs are flexible and efficient to model walking behavior, which corresponds to the operational level in the proposed hierarchy. Route choice, activity choice and destination choice are captured by higher level decision processes. The DCM framework has already been used in similar contexts, in transportation science. A possibility is represented by the extension of the DCM framework from the operational to the tactical and strategical levels. It would allow for a general model of pedestrian behavior, using the same mathematical framework.

- The proposed research makes use of pure dynamic data, where only kinematic characteristics of decision makers are available. Complex behavioral patterns can be captured taking into account individual characteristics. Under controlled experimental conditions, it would be possible to collect individual information, such as the age, sex, weight, destination, level of stress, nature of the trips, area knowledge. DCMs provide a well-founded statistical framework to integrate such socio-economic characteristics, in order to calibrate more detailed and realistic models.

- In the proposed research, only pedestrian-pedestrian interactions have been taken into account. It would be interesting to include the influence of the spatial layout as well as fixed and moving obstacles.

- The detection/tracking system proposed here is based on a correlation-based tracker. More advanced methods proposed in the computer vision literature should be tested, combined with the proposed behavioral model, in order to maximize the performances of the system.

- Most of the vision-based systems installed in public spaces for video-surveillance purposes have to face the problem of event detection. Especially the identification and recognition of *abnormal* situations, such as those dictated by panic and emergency conditions, are crucial. An extension of the proposed model to these scenarios would be useful, in other to integrate behavioral information with video processing techniques, for automatic video surveillance, activity recognition and event detection applications.

Bibliography

Abbé, E., Bierlaire, M. and Toledo, T. (2005). Normalization and correlation of cross-nested logit models, *Technical Report RO-050912*, Operations Research Group ROSO, Institute of Mathematics, Ecole Polytechnique Fédérale de Lausanne.

Agarwal, P. K., Arge, L. and Erickson, J. (2000). Indexing moving points, *Proc. of the 19th ACM Symp. on Principles of Database Systems (PODS)*, pp. 175–186.

Agrawal, R., Lin, K. I. and Sawhney, H. S. (1995). Fast similarity search in the presence of noise, scaling, and translation in times-series databases, *VLDB*.

Ahmed, K. I. (1999). *Modeling drivers' acceleration and lane changing behaviors.*, PhD thesis, Massachusetts Institute of Technology.

Anderberg, M. R. (1973). *Cluster Analysis for Applications*, Academic Press, Inc., New York, NY.

Antonini, G. and Bierlaire, M. (2005). Capturing interactions in pedestrian walking behavior in a discrete choice framework, *Technical Report RO-050921*, Operations Research Group ROSO, Institute of Mathematics, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland.

Antonini, G., Bierlaire, M. and Weber, M. (2004a). Discrete choice models of pedestrian walking behavior, Accepted for publication in *Transportation Research Part B*.

Antonini, G., Bierlaire, M. and Weber, M. (2004b). Simulation of pedestrian behaviour using a discrete choice model calibrated on actual motion data, 4th Swiss Transport Research Conference. www.strc.ch.

Antonini, G. and Thiran, J. P. (2004). Trajectories clustering in ica space: an application to automatic counting of pedestrians in video sequences, *in* J. Blanc-Talon and D. Popescu (eds), *Advanced Concepts for Intelligent Vision Systems (ACIVS) 2004*, Brussels, Belgium.

Antonini, G., Venegas, S., Bierlaire, M. and Thiran, J. P. (2005). Behavioral priors for detection and tracking of pedestrians in video sequences, To appear in International Journal of Computer Vision.

Antoniou, C., Ben-Akiva, M., Bierlaire, M. and Mishalani, R. (1997). Demand simulation for dynamic traffic assignment, *Proceedings of the 8th IFAC Symposium on Transportation Systems*, Chania, Greece. its.mit.edu/pubs/pretrip.ps.

Arentze, T. A. and Timmermans, H. J. P. (2004). Capturing the role of awareness and information search processes on choice set formation in models of activity-travel behavior, *Transportation Research Board, Washington* .

Arulampalam, M. S., Maskell, S., Gordon, N. and Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking, *IEEE Trans.on Signal Processing* **50**(2): 174–188.

Bachelder, I. A. and Waxman, A. M. (1994). Mobile robot visual mapping and localization: a view-based neurocomputational architecture that emulates hippocampal place learning, *Neural Networks* **7**(6/7): 1083–1099.

Balasubramanian, M., Shwartz, E. L., Tenenbaum, J. B., de Silva, V. and Langford, J. C. (2002). The isomap algorithm and topological stability, *Science* **295**(5552).

Barceló, J. and Ferrer, J. L. (1997). *AIMSUN2: Advanced Interactive Microscopic Simulator for Urban Network*, Departement d'Estadística i Investigació Operativa, Universitat Politècnica de Catalunya. User's Manual.

Bartlett, M., Movellan, J. and Sejnowski, T. (2002). Face recognition by independent component analysis, *IEEE Transactions on neural networks* **13**(6): 1450–1464.

Bath, C. R. and Guo, J. (2004). A mixed spatially correlated logit model: Formulation and application to residential choice modeling, *Transportation Research Part B* **38**(2): 147–168.

Bell, A. J. and Sejnowski, T. J. (1995). An information maximisation approach to blind separation and blind deconvolution, *Neural Computation* **7**(6): 1129–1159.

Ben-Akiva, M. and Bierlaire, M. (1999). Discrete choice methods and their applications to short-term travel decisions, *in* R. Hall (ed.), *Handbook of Transportation Science*, Kluwer, pp. 5–34.

Ben-Akiva, M., Bierlaire, M., Koutsopoulos, H. N. and Mishalani, R. (2002). Real-time simulation of traffic demand-supply interactions within DynaMIT, *in* M. Gendreau and P. Marcotte (eds), *Transportation and network analysis: current trends. Miscellenea in honor of Michael Florian*, Kluwer Academic Publishers, Boston/Dordrecht/London.

Ben-Akiva, M. and Bolduc, D. (1996). Multinomial probit with a logit kernel and a general parametric specification of the covariance structure. Working Paper, Department of Civil Engineering, MIT.

Ben-Akiva, M. E., Bergman, M. J., Daly, A. J. and Ramaswamy, R. (1984). Modeling inter-urban route choice behaviour, *in* J. Volmuller and R. Hamerslag (eds), *Proceedings from the ninth international symposium on transportation and traffic theory*, VNU Science Press, Utrecht, Netherlands, pp. 299–330.

Ben-Akiva, M. E. and Lerman, S. R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, Cambridge, Ma.

Berbdt, D. and Clifford, J. (1994). Using dynamic time warping to find patterns in time series, *In Proc. of KDD Workshop*.

Beymer, D. and Konolige, K. (1999). Real-time tracking of multiple people using continous detection.

Bezdek, J. C. (1981). *Pattern Recognition With Fuzzy Objective Function Algorithms.*, Plenum Press, New York, NY.

Bierlaire, M. (2003). BIOGEME: a free package for the estimation of discrete choice models, *Proceedings of the 3rd Swiss Transportation Research Conference*, Ascona, Switzerland. www.strc.ch.

Bierlaire, M. (2005). Capturing complex transportation behaviors with advanced random utility models, *in* H. van Zuylen (ed.), *The reliability of traveling and the robustness of transport systems*, TRAIL Research School.

Bierlaire, M., Antonini, G. and Weber, M. (2003). Behavioral dynamics for pedestrians, Moving through nets: the physical and social dimensions of travel. In Press.

Bierlaire, M., Mishalani, R. and Ben-Akiva, M. (2000). General framework for dynamic demand simulation, *Technical Report RO-000223*, ROSO-DMA-EPFL Swiss Institute of Technology, CH-1015 Lausanne. rosowww.epfl.ch/mbi/demand-report.pdf.

Biliotti, D., Antonini, G. and Thiran, J. P. (2005). Multi-layer hierarchical clustering of pedestrian trajectories for automatic counting of people in video sequences, *IEEE Motion 2005*, Beckenridge, Colorado, USA.

Blue, V. J. and Adler, J. L. (2001). Cellular automata microsimulation for modeling bi-directional pedestrian walkways, *Transportation Research Part B* **35**(3): 293–312.

Blue, V. J. and Adler, J. L. (2002). Flow capacities from cellular automata modeling of proportional splits of pedestrians by direction, *in* M. Schreckenberg and S. Sharma (eds), *Pedestrian and Evacuation Dynamics*, Springer, pp. 115–122.

Boley, D. (1998). Principal direction divisive partitioning, *Data Mining and Knowledge Discovery* **2**(4): 325–344.

Borgers, A. and Timmermans, H. (1986a). City centre entry points, store location patterns and pedestrian route choice behaviour: A micro-level simulation model, *Socio-Economic Planning Sciences* **20**(1): 25–31.

Borgers, A. and Timmermans, H. (1986b). A model of pedestrian route choice and demand for retail facilities within inner-city shopping areas, *Geographical analysis* **18**(2): 115–128.

Bovy, P. H. L. and Stern, E. (1990). *Route choice: wayfinding in transport networks*, Vol. 9 of *Studies in Operational Regional Science*, Kluwer Academic Publishers, Dordrecht, NL.

Box, G. P. and Jenkins, G. M. (1970). *Time series analysis, forecasting and control*, Holden-Day, San Francisco, Ca.

Brady, A. T. and Walker, M. B. (1978). Interpersonal distance as a function of situationally induced anxiety, *British Journal of Social and Clinical Psychology* **17**: 127–133.

Bregler, C. (1997). Learning and recognizing human dynamics in video sequences, *CVPR '97: Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*, IEEE Computer Society, Washington, DC, USA, p. 568.

Cai, Q. and Aggarwal, J. (1996). Tracking human motion using multiple cameras, *Proc. of International Conference on Pattern Recognition*, Vienna, pp. 68–72.

Cascetta, E., Nuzzolo, A. and Biggiero, L. (1992). Analysis and modeling of commuters' departure time and route choice in urban networks, *Proceedings of the second international Capri seminar on urban traffic networks*.

Charlesworth, J. A. and Gunawan, E. (1987). The behavioural dynamics of route choice, *Traffic Engineering and Control* pp. 80–83.

Chernew, M., Gowrisankaran, G. and Scanlon, D. (2003). Learning and the value of information: The case of health plan report cards, *International Economic Review, submitted to* .

Clason, R. (1990). Finding clusters: an application of the distance concept., *The Mathematic Teacher* .

Collins, R., Lipton, A., Kanade, T., Fujiyoshi, H., Duggins, D., Tsin, Y., Tolliver, D., Enomoto, N. and Hasegawa, O. (2000). A system for video surveillance and monitoring, *Technical Report CMU-RI-TR-00-12*, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA.

Comaniciu, D., Ramesh, V. and Meer, P. (2003). Kernel-based object tracking, *IEEE Trans. on Pattern Analysis and Machine Intelligence* **25**(5): 564–577.

Conroy, R. A. (2001). *Spatial Navigation in Immersive Virtual Environments*, PhD thesis, University of London.

Costella, J. P. (1992). *Galilean Antialiasing for Virtual Reality Displays*, sci.virtual-worlds, University of Melbourne.

Cutler, R. and Davis, L. S. (2000). Robust real-time periodic motion detection, analysis and applications, *IEEE Trans. Patt. An. Mach. Int.* **22**(8): 781–796.

Daamen, W. (2004). *Modelling Passenger Flows in Public Transport Facilities*, PhD thesis, Delft University of Technology.

Dave, R. N. (1992). Generalized fuzzy c-shells clustering and detection of circular and elliptic boundaries., *Pattern Recogn.* **25**: 713–722.

DeCarlo, D. and Metaxas, D. (2000). Optical flow constraints on deformable models with applications to face tracking, *Int.J.of Computer Vision* **38**(2): 99– 127.

Dellaert, B. G. C., Arentze, T. A., Bierlaire, M., Borgers, A. W. and Timmermans, H. J. (1998). Investigating consumers' tendency to combine multiple shopping purposes and destinations, *Journal of Marketing Research* **35**(2): 177–188.

Dijkstra, J., Timmermans, H. J. P. and Jessurun, A. J. (2000). A multi-agent cellular automata system for visualising simulated pedestrian activity, *in* S. Bandini and T. Worsch (eds), *Theoretical and Practical Issues on Cellular Automata - Proceedings on the 4th International Conference on Cellular Automata for research and Industry*, Springer Verlag.

Dosey, M. A. and Meisels, M. (1969). Personal space and self-protection, *Journal of Personality and Social Psychology* **11**(2): 93–97.

Dubes, R. C. and Jain, A. K. (1976). Clustering techniques: The user dilemma, *Pattern Recognition* **8**: 247–260.

Efros, A., Berg, A. C., Mori, G. and Malik, J. (2003). Recognizing action at a distance, *IEEE International Conference on Computer Vision*, Nice, France, pp. 726–733.

Eiter, T. and Mannila, H. (1997). Distance measures for point sets and their computation, *Acta Informatica* **34**(2): 109–133.

Elgammal, A., Duraiswami, R., Harwood, D. and Davis, L. S. (2002). Background and foreground modeling using nonparametric kernel density estimation for visual surveillance, *Proc. IEEE* **90**(7): 1151–1163.

Ferber, J. (1998). *Multi-agent systems*, Addison-Wesley, Reading, Ma.

Foster, J. P., Nixon, M. S. and Bennett, A. (2003). Automatic gait recognition using area-based metrics, *Pattern Recogn. Lett.* **24**(14): 2489–2497.

Fuerstenberg, K. C., Dietmayer, K. C. J. and Willhoeft, V. (2002). Pedestrian recognition in urban traffic using a vehicle based multilayer laserscanner, *Proceedings of IEEE Intelligent Vehicles Symposium*, Versailles, France.

Fujiyoshi, H. and Lipton, A. (1998). Real-time human motion analysis by image skeletonisation, *Proc. IEEE WACV 98*, pp. 15–21.

Gavrila, D. M. (1999). The visual analysis of human movement: A survey, *Computer Vision and Image Understanding: CVIU* **73**(1): 82–98.

Golledge, R. G. (1993). Geographical Perspectives on Spatial Cognition, *in* T. Garling and R. G.Golledge (eds), *Behavior and Environment:Psychological and Geographical Approaches*, Elsevier Science.

Golledge, R. G. (1999). *Wayfinding Behavior:Cognitive Mapping and Other Spatial Processes*, Johns Hopkins University Press, Baltimore, MD.

Golledge, R. G. (2002). Dynamics and ITS: behavioural responses to information available from ATIS, *in* H. Mahmassani (ed.), *Perpetual motion: travel behaviour research opportunities and application challenges*, Elsevier Science.

Guo, B., Lam, K.-M., Lin, K.-H. and Siu, W.-C. (2003). Human face recognition based on spatially weighted hausdorff distance, *Pattern Recogn. Lett.* **24**(1-3): 499–507.

Gwynne, S., Galea, E. R., Lawrence, P., Owen, M. and Filippidis, L. (1997). Exodus. http://fseg.gre.ac.uk/exodus/.

Han, B., Comaniciu, D., Zhu, Y. and Davis, L. (2004). Incremental density approximation and kernel-based bayesian filtering for object tracking, *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 638–644.

Haritaoglu, I. (1998). A real-time system for detection and tracking of people and recognizing their activities, PhD(proposal).

Hartnett, J. J., Bailey, K. G. and Hartley, C. S. (1974). Body height, position, and sex as determinants of personal space, *Journal of Psychology* **87**: 129–136.

Heikkila, J. (2000). Geometric camera calibration using circular control points, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(10): 1066–1077.

Heisele, B., Kressel, U. and Ritter, W. (1997). Tracking non-rigid, moving objects based on color cluster flow, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, San Juan, pp. 257–260.

Helbing, D., Farkas, I. J., Molnar, P. and Vicsek, T. (2002). Simulation of pedestrian crowds in normal and evacuation simulations, *in* M. Schreckenberg and S. Sharma (eds), *Pedestrian and Evacuation Dynamics*, Springer, pp. 21–58.

Helbing, D., Farkas, I. and Vicsek, T. (2000). Simulating dynamical features of escape panic, *Nature* **407**(28): 487–490.

Helbing, D. and Molnár, P. (1995). Social force model for pedestrian dynamics, *Physical review E* **51**(5): 4282–4286.

Hensher, D. A. and Johnson, L. W. (1981). *Applied Discrete-Choice Modelling*, Croom Helm, London.

Hensher, D. and Greene, W. (2001). The mixed logit model: The state of practice and warnings for the unwary. Working Paper, University of Sydney.

Herman, R. and Rothery, R. W. (1965). Car following and steady-state flow, *Proceedings on 2nd international symposium on the theory of traffic flow*, pp. 1–11.

Hess, S., Bierlaire, M. and Polak, J. W. (2005). Capturing taste heterogeneity and correlation structure with mixed gev models, *84th Annual Meeting of the Transportation Research Board*.

Hillier, B. and Hanson, J. (1984). *The Social Logic of Space*, Cambridge University Press, Cambridge.

Hillier, B., Penn, A., Hanson, J. and Grajewski, T. (1993). Natural movement;or, configuration and attraction in urban space use, *Environment and Planning B: Planning and Design* **20**: 29–66.

Hoogendoorn, S. P. (2003). Pedestrian travel behavior modeling, *10th International Conference on Travel Behavior Research*, Lucerne.

Hoogendoorn, S. P., Bovy, P. H. L. and Daamen, W. (2002). Microscopic pedestrian wayfinding and dynamics modelling, *in* M. Schreckenberg and S. Sharma (eds), *Pedestrian and Evacuation Dynamics*, Springer, pp. 123–155.

Horowitz, J. J., Duff, D. F. and Stratton, L. O. (1964). Body buffer zone:exploration of personal space, Archives of General Psychiatry. 11(6).

Hyvärinen, A. and Oja, E. (2000). Independent component analysis: Algorithms and applications., *Neural Networks* **13**(4-5): 411–430.

Isard, M. and Blake, A. (1996). Contour tracking by stochastic propagation of conditional density, *ECCV* **1**: 343–356.

Isard, M. and Blake, A. (1998). Condensation - conditional density propagation for visual tracking, *International Journal on Computer Vision* **1**(29): 5–28.

Jabri, S., Duric, Z., Wechsel, H. and Rosenfeld, A. (2000). Detection and localisation of people in video images using adaptive fusion of color and edge information, *in* Fairfax (ed.), *Proc. IEEE Int. Conf. Patt. Rec.*, USA, pp. 627–630.

Jain, A. K. and Dubes, R. C. (1988). *Algorithms for Clustering Data.*, Prentice-Hall Inc., Upper Saddle River, NJ.

Jiang, B. (1999). SimPed: simulating pedestrian flows in a virtual urban environment, *Journal of geographic information and decision analysis* **3**(1): 21–30.

Jurie, F. and Dhome, M. (2001). Real time 3d template matching, *International Conference on Computer Vision and Pattern Recognition*, Hawai, pp. 791–797.

Kahn, R. E., Swain, M. J., Prokopowicz, P. N. and Firby, R. J. (1996). Gesture recognition using the perseus architecture, *Proceedings of the 1996 Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, Washington, DC, USA, p. 734.

Kaneko, T. and Hori, O. (2003). Feature selection for reliable tracking using template matching, *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 796–802.

Keogh, E. and Pazzani, M. (2000). Scaling up dynamic time warping for datamining applications, *In Proc. 6th Int.Conf. on Knowledge Discovery and Data Mining*, Boston, MA.

Kessel, A., Klüpfel, H., Wahle, J. and Schreckenberg, M. (2002). Microscopic simulation of pedestrian crowd motion, *in* M. Schreckenberg and S. Sharma (eds), *Pedestrian and Evacuation Dynamics*, Springer, pp. 193–200.

King, B. (1967). Step-wise clustering procedures., *J. Am. Stat. Assoc.* **69**: 86–101.

Kitagawa, G. (1996). Monte carlo filter and smoother for non-gaussian nonlinear state space models, *Journal of Computational and Graphical Statistics* **5**(1): 1–25.

Klüpfel, H., Meyer-König, M., Wahle, J. and Schreckenberg, M. (2000). Microscopic simulation of evacuation processes on passenger ships, *in* S. Bandini and T. Worsch (eds), *Theoretical and Practical Issues on Cellular Automata*, London, pp. 63–71.

Koning, R. H. (1991). Discrete choice and stochastic utility maximization, Research Memorandum 414, Department of Economics, Groningen University.

Koning, R. H. and Ridder, G. (1994). On the compatibility of nested logit models with utility maximization, *Journal of Econometrics* **63**: 389–396.

Kopp, N. (1999). EvacSim, an evacuation simulation model. www.nathan.kopp.com/EvacSim/evacsim.html.

Lee, G. (1966). A generalization of linear car following theory, *Operation Research 14* pp. 595–606.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals, Soviet Physics Doklady, 10(8) S. 707-710.

Lombardi, P. (2001). A survey on pedestrian detection for autonomous driving systems, Technical Report.

Luck, J. P., Debrunner, C., Hoff, W., He, Q. and Small, D. E. (2002). Development and analysis of a real-time human motion tracking system, *WACV 2002*, pp. 196–202.

Maggio, E. and Cavallaro, A. (2005). Hybrid particle filter and mean shift tracker with adaptive transition model, *Proc. of IEEE Signal Processing Society International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Philadelphia, PA, USA, pp. 19–23.

Mahmassani, H., Hu, T. Y., Peeta, S. and Ziliaskopoulos, A. (1993). Development and testing of dynamic traffic assignment and simulation procedures for ATIS/ATMS applications, *Technical Report DTFH61-90-R-00074-FG*, Center for Transportation Research, University of Texas at Austin.

Mahmassani, H. S. (1996). Dynamics of commuter behaviour: recent research and continuing challenges, *in* P. Stopher and M. Lee-Gosselin (eds), *Understanding Travel Behaviour in an Era of Change*, Elsevier, pp. 279–313.

Manski, C. (1977). The structure of random utility models, *Theory and Decision* **8**: 229–254.

Manski, C. F. and McFadden, D. (1981). *Econometric models of probabilistic choice*, in C.F. Manski and D. McFadden, editors, Structural Analysis of Discrete Data with Econometric Applications, MIT Press, Cambridge, 198-272.

Mao, J. and Jain, A. K. (1996). A self-organizing network for hyperellipsoidal clustering (hec)., *IEEE Trans. Neural Netw.* **7**: 16–29.

McFadden, D. (1978). Modelling the choice of residential location, *in* A. Karlquist *et al.* (ed.), *Spatial interaction theory and residential location*, North-Holland, Amsterdam, pp. 75–96.

McFadden, D. and Train, K. (2000). Mixed MNL models for discrete response, *Journal of Applied Econometrics* **15**(5): 447–470.

Mitchell, T. (1997). *Machine Learning.*, McGraw- Hill, Inc., New York, NY.

Muramatsu, M., Irie, T. and Nagatani, T. (1999). Jamming transition in pedestrian counter flow, *Physica A* **267**: 487–498.

Nagel, K. (in progress). *Multi-agent transportation simulations.* http://www.vsp.tu-berlin.de/publications/matsim-book/.

Newell, G. (1961). Nonlinear effects in the dynamics of car following, *Operation Research 9* pp. 209–229.

Nummiaro, K., Koller-Meier, E. and Gool, L. V. (2002). Object tracking with an adaptive color-based particle filter, *in* L. V. Gool (ed.), *Symposium for Pattern Recognition of the DAGM*, Springer, pp. 353–360.

Nummiaro, K., Koller-Meier, E., Svoboda, T., Roth, D. and Gool, L. V. (2003). Color-based object tracking in multi-camera environments, *in* B. Michaelis and G. Krell (eds), *25th Pattern Recognition Symposium, DAGM 2003*, LNCS, Springer, pp. 591–599.

Okazaki, S. (1979). A study of pedestrian movement in architectural space. Part 1: pedestrian movement by the application of magnetic models, *Trans. of A.I.J.* **283**: 111–119.

Oliver, N. M., Rosario, B. and Pentland, A. P. (2000). A bayesian computer vision system for modeling human interactions, *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(8): 831–843.

Papageorgiou, C. and Poggio, T. (1999). A pattern classification approach to dynamical object detection, *ICCV '99: Proceedings of the International Conference on Computer Vision-Volume 2*, IEEE Computer Society, Washington, DC, USA, p. 1223.

Paragios, N. and Tziritas, G. (1999). Adaptive detection and localization of moving objects in image sequences, *Signal Process.: Image Commun.* **14**(4): 277–296.

Parker, D. C., Manson, S. M., Janssen, M. A., Hoffmann, M. J. and Deadman, P. (forthcoming). Multi-agent systems for the simulation of land-use and land-cover change: a review, *Annals of the Association of American Geographers* . www.csiss.org/events/other/agent-based/papers/maslucc_overview.pdf.

Penn, A. (2003). Vision, configuration and simulation of static interaction for design, *in* E.Galea (ed.), *Pedestrian and Evacuation Dynamics*, CMS Press, University of Greenwich, London.

Penn, A. and Turner, A. (2002). Space syntax based agent simulation, *in* M. Schreckenberg and S. Sharma (eds), *Pedestrian and Evacuation Dynamics*, Springer, pp. 99–114.

Perng, S., Wang, H., Zhang, S. and Parker, D. S. (2000). Landmarks:a new model for similarity-based pattern querying in time series databases, *In Proceedings if ICDE*, pp. 33–42.

Phillips, J. R. (1979). An exploration of perception of body boundary, personal space, and body size in elderly persons, *Perceptual and Motor Skills* **48**: 299–308.

Polana, R. and Nelson, R. C. (1997). Detection and recognition of periodic, nonrigid motion, *Int. J. Comput. Vision* **23**(3): 261–282.

Quarantelli, E. L. (2001). Sociology of panic, *International Encyclopedia of the Social & Behavioral Sciences*, Elsevier Science.

Ramming, M. S. (2001). *Network Knowledge and Route Choice*, PhD thesis, Massachusetts Institute of Technology.

Ramon, J. and Bruynooghe, M. (2001). A polynomial time computable metric between point sets, *Acta Informatica* **37**(10): 765–780.

Reynolds, C. W. (1987). Flocks, herds, and schools: A distributed behavioral model, *Computer Graphics* **21**(4): 25–34.

Rickert, M., Nagel, K., Schreckenberg, M. and Latour, A. (1996). Two lane traffic simulations using cellular automata, *Physica A* **231**: 534–550.

Sanders, J. L. (1976). Relationship of personal space to body image boundary definiteness, *Journal of Research in Personality* **10**: 478–481.

Saul, S. R. . L. (2000). Nonlinear dimensionality reduction by locally linear embedding, *Science* **290**(5500): 2323–2326.

Schadschneider, A. (2002). Cellular automaton approach to pedestrian dynamics — Theory, *in* M. Schreckenberg and S. Sharma (eds), *Pedestrian and Evacuation Dynamics*, Springer, pp. 75–86.

Schadschneider, A., Kirchner, A. and Nishinari, K. (2002). CA approach to collective phenomena in pedestrian dynamics, *in* S. Bandini, B. Chopard and M. Tomassini (eds), *Cellular Automata, 5th International Conference on Cellular Automata for Research and Industry, ACRI 2002, Geneva, Switzerland, October 9-11, 2002, Proceedings*, Vol. 2493 of *Lecture Notes in Computer Science*, Springer, pp. 239–248.

Schreckenberg, M. and Sharma, S. (eds) (2002). *Pedestrian and Evacuation Dynamics*, Springer Verlag.

Schultz, D. (ed.) (1964). *Panic behavior: discussion and readings*, Random House, New-York.

Senior, A. W. (2002). Tracking with probabilistic appearance models, *Proc. ECCV workshop on Performance Evaluation of Tracking and Surveillance Systems*, pp. 48–55.

Shechtman, E. and Irani, M. (2005). Space-time behavior based correlation, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 405–412.

Shio, A. and Sklansky, J. (1991). Segmentation of people in motion, *IEEE Workshop on Visual Motion*, pp. 325–332.

Sminchisescu, C. and Triggs, B. (2003). Estimating articulated human motion with covariance scaled sampling, *International Journal of Robotics Research* **22**: 371–391.

Sommer, R. (1969). *Personal Space: The behavioral bases of design*, Prentice Hall, Englewood Cliffs, NJ.

Souvenir, R. and Pless, R. (2005). Isomap and nonparametric models of image deformation, *Motion05*, pp. 195–200.

Stauffer, C. and Grimson, W. E. L. (2000). Learning patterns of activity using real-time tracking, *IEEE Trans. Pattern Anal.Mach.Intell.* **22**: 747–757.

Stone, J. V. (2002). Independent component analysis: An introduction., *Trends in Cognitive Sciences* **6**(2): 59–64.

Strehl, A. and Ghosh, J. (2000). Value-based customer grouping from large retail datasets, *Proc. of the SPIE Conf. on Data Mining and Knowledge Discovery*, Orlando.

Symon, M. J. (1977). Clustering criterion and multi-variate normal mixture., *Biometrics* pp. 35–43.

Teknomo, K. (2002). *Microscopic Pedestrian Flow Characteristics: Development of an Image Processing Data Collection and Simulation Model*, PhD thesis, Tohoku University, Japan, Sendai.

Teknomo, K., Takeyama, Y. and Inamura, H. (2000). Review on microscopic pedestrian simulation model, *Proceedings Japan Society of Civil Engineering Conference*.

Teknomo, K., Takeyama, Y. and Inamura, H. (2001). Microscopic pedestrian simulation model to evaluate "lane-like segregation" of pedestrian crossing, *Proceedings of Infrastructure Planning Conference*, Vol. 24.

Terzopoulos, D., Witkin, A. and Kass, M. (1988). Constraints on deformable models: Recovering 3d shape and nonrigid motion, *Artificial Intelligence* **36**(1): 91–123.

Thalmann, D. and Bandi, S. (1998). Space discretization for efficient human navigation, *Eurographics '98*, Computer Graphics Forum, pp. 195–206.

Thayananthan, A., Stenger, B., Torr, P. H. S. and Cipolla, R. (2003). Learning a kinematic prior for tree-based filtering, *Proc. British Machine Vision Conference*, Vol. 2, Norwich, UK, pp. 589–598.

Thompson, P. A. and Marchant, E. W. (1994). Simulex. http://www.ies4d.com/.

Toffoli, T. and Margolus, N. (1987). *Cellular Automata Machines*, MIT Press, Cambridge, Ma.

Toledo, T. (2003). *Integrated Driving Behavior Modeling.*, PhD thesis, Massachusetts Institute of Technology.

Train, K. (2003). *Discrete Choice Methods with Simulation*, Cambridge University Press, University of California, Berkeley.

Turner, A. (2001). Angular analysis, *In Proceedings 3rd International Symposium on Space Syntax*, pp. 30.1–30.11.

Turner, A., Doxa, M., O'Sullivan, D. and Penn, A. (2001). From isovists to visibility graphs: a methodology for the analysis of architectural space, *Environment and Planning B* **28**(1): 103–121.

Turner and Penn (2002). Evas. http://www.vr.ucl.ac.uk/research/evas/.

Vlachos, M., Kollios, G. and Gunopulos, D. (2002). Discovering similar multidimensional trajectories, *ICDE*.

Walker, J. L. (2001). *Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables*, PhD thesis, Massachusetts Institute of Technology.

Wallace, M. and Kollias, S. (2004). Robust, generalized, quick efficient agglomerative clustering, *Proc. of 6th Intern. Conf. on Enterprise Inf. Syst.*, Porto, Portugal.

Webb, J. D. and Weber, M. J. (2003). Influence of sensor abilities on the interpersonal distance of the elderly, *Environment and behavior* **35**(5): 695–711.

Wen, C. H. and Koppelman, F. S. (2001). The generalized nested logit model, *Transportation Research B* **35**(7): 627–641.

Whynes, D. K., Reedand, G. and Newbold, P. (1996). General practitioners' choice of referral destination: A probit analysis, *Managerial and Decision Economics* **17**(6): 587.

Willson, G. (1994). Modeling and calibration of automated zoom lenses, *Proceedings of the SPIE 2350:Videometrics III*.

Wren, C., Azarbayejani, A., Darrell, T. and Pentland, A. (1997). Pfinder: Realtime tracking of the human body, *IEEE Trans. Pattern Anal.Mach.Intell.* **19**(7): 780–785.

Yang, L., Fang, W., Huang, R. and Deng, Z. (2002). Occupant evacuation model based on cellular automata in fire, *Chinese Science Bulletin* **47**(17): 1484–1488.

Yang, Q. and Koutsopoulos, H. (1997). A microscopic traffic simulator for evaluation of dynamic traffic management systems, *Transportation Research Part C* **4**(3): 113–129.

Yi, B.-K. and Faloutsos, C. (2000). Fast time sequence indexing for arbitrary lp norms, *Proceedings of the 26th International Conference on Very Large Data Bases*, Morgan Kaufmann Publishers Inc., pp. 385–394.

Zadeh, L. A. (1965). Fuzzy sets, *Inf. Control* **8**: 338–353.

Gianluca Antonini

Chemin du Croset 11
1024, Ecublens, Switzerland
gianluca.antonini@epfl.ch
gianluca@mit.edu
+41216915119
+41788198330

Educational background

Massachusetts Institute of Technology, Boston US (2004 / 2005)
Visiting PhD Candidate - Temporary position: Working on "Discrete Choice Models and modeling of pedestrian behavior", Dept. of Civil Engineering – Centre for Intelligent Transportation System
Course work preparation for Demand Modeling, and Discrete Choice Models for Prof. Moshe Ben Akiva

Ecole Polytechnique Fédérale de Lausanne (EPFL) (01/2003 - present)
Signal Processing Institute STI/ITS
Ph.D. Candidate - Project: "Behavioral model-based scene analysis"
Defense date: 11/2005
Advisors: Prof. Jean Philippe Thiran (STI/ITS) and Dr. Michel Bierlaire (Mathematics Institute)
Institute Director: Prof. Murat Kunt

University of Siena, Siena, Italy (Academic years: 1992/1993 - 1999/2000)
Master degree in Telecommunication Engineering (108/110), Dept. of Information Engineering
Master project: "Video segmentation and key frame extraction"
Advisors: Prof. Alessandro Mecocci

Technical Institute "A.Manetti" (09/1987 - 06/1992)
High school degree in technical studies (58/60)

Advisor for Master degree projects

R.Ceccarelli, G.Antonini and J.P.Thiran
Pedestrian head detection using automatic scale selection for feature detection and statistical edge curvature analysis
September 2004

D.Biliotti, G.Antonini and J.P.Thiran
Multi-layer hierarchical clustering of pedestrian trajectories for automatic counting people in video sequences
September 2004

Professional experience

Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland (2003)
Signal Processing Institute ITS/STI
Project: SIMBAD: SIMulation Based Automatic Detection
In collaboration with Visiowave SA

Ecole Polytechnique Federale de Lausanne (EPFL), Switzerland (06/2002 - 01/2003)
Signal Processing Institute ITS/STI
Research Assistant - Project: BANCA: Facial feature extraction using Support Vector Machine and Independent Component Analysis

Fastcom Technology SA, Lausanne, Switzerland (03/2001 - 03/2002)
R&D Engineer: Project: iMVS cameras for Intelligent Machine Vision Systems

Teaching experience

Individual choice behavior : Theory and Application of Discrete Choice Analysis
Workbook of case studies on the application of discrete choice methods in marketing and transportation science. Used at MIT Professional Institute Summer Courses and at EPFL Advanced Continuing Education Course.

Relevant courses to the PhD program

Institute for Pure and Applied Mathematics, UCLA, Los Angeles
Intelligent Extraction of Information from Graphs and High Dimensional Data.
Organizing committee: Kevin Vixie, Chair (Los Alamos National Laboratory), Edmond Chow (D.E. Shaw Research & Development), Tina Eliassi-Rad (Lawrence Livermore National Laboratory), Yann LeCun (New York University), Carey Priebe (Johns Hopkins University)

Operation Research Group, ROSO, EPFL, Lausanne, Switzerland
Discrete Choice Analysis: Predicting Demand and Market Shares
Lecturers: Prof. Moshe Ben Akiva**,** *Massachusetts Institute of Technology,* Dr. Michel Bierlaire, *Ecole Polytechnique Fédérale de Lausanne,* Prof. Denis Bolduc, *University Laval, Québec,* Prof. Daniel McFadden**,** *University of California, Berkeley* (Nobel Prize Laureate, 2000)

Fellowships and awards

Granted fellowship for participation at the “Intelligent Extraction of Information from Graphs and High Dimensional Data” graduate summer school, 2005

Full scholarship for MIT Visiting Program, 2004 - 2005

Granted fellowship for participation at the “Discrete Choice Analysis: Predicting Demand and Market Shares” course, 2003

Full scholarship for PhD program at STI/ITS, EPFL, Lausanne, Switzerland, 2002 – 2005

Informatics skills

Operating systems: Windows platforms, basic Linux

Programming languages: C/C++, Matlab, basics of Java

Language skills

Italian: Mother Tongue
English: Fluent (oral and written)
French: Fluent (oral), intermediate (written)
Spanish: Basic understanding

Other research interests

Behavioral models, decision theory, econometric, machine learning, artificial intelligence, data analysis, image processing

Other information

Playing electric guitar since 1987
Jazz studies with Mr. Umberto Fiorentino in Rome, Italy, and at EJMA Music Academy in Lausanne, Switzerland

List of publications

Book chapters:

S.Venegas, G.Antonini, J.P.Thiran and M.Bierlaire (2005). **Automatic Pedestrian Tracking Using Discrete Choice Models and Image Correlation Techniques**, *in* Bengio and Bourlard (eds), *Lecture Notes in Computer Science*, Volume 3361, pp. 341 – 348

G.Antonini, V.Popovici and J.P.Thiran
Independent Component Analysis and Support Vector Machine for Face Feature Extraction
Lecture Notes in Computer Science, 4th International Conference on Audio- and Video-Based Biometric Person Authentication, Guildford, UK, Vol. 2688, pp. 111-118, June 2003

Journal papers:

G.Antonini, S.Venegas, M.Bierlaire and J.P.Thiran
Behavioral priors for detection and tracking of pedestrians in video sequences
International Journal of Computer Vision, 2004 (in press)

G.Antonini, M.Bierlaire and M.Weber
Discrete Choice Models of Pedestrian Walking Behavior
Transportation Research Part B, 2004 (in press)

Conference papers:

G.Antonini and M.Bierlaire
A Discrete choice framework for acceleration and direction change behaviors in walking pedestrians
Proceedings of Pedestrian and Evacuation Dynamics, September 2005, Vienna, Austria

D.Biliotti, G.Antonini and J.P.Thiran
Multi-layer hierarchical clustering of pedestrian trajectories for automatic counting of people in video sequences
Proceedings of the IEEE Workshop on Motion and Video Computing (WACV/MOTION/PETS), January 2005

S.Venegas, G.Antonini, J.P.Thiran and M.Bierlaire
Bayesian Integration of a Discrete Choice Pedestrian Behavioral Model and Image Correlation Techniques for Automatic Multi Object Tracking
2004 IEEE International Conference on Image Processing, ICIP2004, Singapore, October 2004

G.Antonini, S.Venegas, J.P.Thiran and M.Bierlaire
A discrete choice pedestrian behavior model for pedestrian detection in visual tracking systems
Advanced Concepts for Intelligent Vision Systems, ACIVS 2004, Brussels, Belgium, September 2004

G.Antonini and J.P.Thiran
Trajectories clustering in ICA space: an application to automatic counting of pedestrians in video sequences
Advanced Concepts for Intelligent Vision Systems, ACIVS 2004, Brussels, Belgium, September 2004

G.Antonini, M.Bierlaire and M.Weber
Simulation of Pedestrian Behavior using a Discrete Choice Model Calibrated on Actual Motion Data
4th Swiss Transport Research Conference, March 2004

M.Bierlaire, G.Antonini and M.Weber
Behavioral dynamics for pedestrians
Moving through nets: the physical and social dimensions of travel, 2003

Technical Reports:

G. Antonini and M. Bierlaire
Capturing interactions in pedestrian walking behavior in a discrete choice framework
Submitted to Transportation Research Part B, September 2005

G.Antonini and J.P.Thiran
Counting pedestrians in video sequences using trajectory clustering: a comparative study
Submitted to Circuits and Systems for Video Technologies, September 2005