Building and Leveraging Prior Knowledge for Predicting Pedestrian Behaviour Around Autonomous Vehicles in Urban Environments

Pavan Vasishta

Communauté
UNIVERSITÉ Grenoble Alpes

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTE UNIVERSITE GRENOBLE ALPES

Spécialité : **Mathematique et Infomatique**

Arrêté ministériel : 25 mai 2016

Présentée par

Pavan VASISHTA

Thèse dirigée par **Anne SPALANZANI**, et
codirigée par **Dominique VAUFREYDAZ**

préparée au sein du **Laboratoire Inria Grenoble**
dans **l'École Doctorale Mathématiques, Sciences et technologies de l'information, Informatique**

Building and Leveraging Prior Knowledge for Predicting Pedestrian Behaviour Around Autonomous Vehicles in Urban Environments

Acquisition et exploitation des connaissances antérieures pour prédire le comportement des piétons autour des véhicules autonomes en environnement urbain

Thèse soutenue publiquement le **30 Septembre 2019**,
devant le jury composé de :

M. Thierry CHATEAU
Professeur, Clermont Auvergne Université, France, Rapporteur
M. Luis MERINO
Professeur, Universidad Pablo de Olivade, Espagne, Rapporteur
M. François CHARPILLET
Directeur de Recherche, Inria Nancy, France, Examinateur
M. Philippe MARTINET
Directeur de Recherche, Inria Nancy, France, Examinateur
M. Didier DONSEZ
Professeur, Université Grenoble Alpes, France, Examinateur, Président
Mme. Anne SPALANZANI
Maître de Conférences, Université Grenoble Alpes, Directrice de thèse
M. Dominique VAUFREYDAZ
Maître de Conférences, Université Grenoble Alpes, Co-Directeur de thèse

Abstract

Autonomous Vehicles navigating in urban areas interact with pedestrians and other shared space users like bicycles throughout their journey either in open areas, like urban city centers, or closed areas, like parking lots. As more and more autonomous vehicles take to the city streets, their ability to understand and predict pedestrian behaviour becomes paramount. This is achieved by learning through continuous observation of the area to drive in. On the other hand, human drivers can instinctively infer pedestrian motion on an urban street even in previously unseen areas. This need for increasing a vehicle's situational awareness to reach parity with human drivers fuels the need for larger and deeper data on pedestrian motion in myriad situations and varying environments.

This thesis focuses on the problem of reducing this dependency on large amounts of data to predict pedestrian motion accurately over an extended horizon. Instead, this work relies on Prior Knowledge, itself derived from the JJ Gibson's sociological principles of "Natural Vision" and "Natural Movement". It assumes that pedestrian behaviour is a function of the built environment and that all motion is directed towards reaching a goal. Knowing this underlying principle, the cost for traversing a scene from a pedestrian's perspective can be divined. As a result, inference on their behaviour can be performed. This work presents a contribution to the framework of understanding pedestrian behaviour as a confluence of probabilistic graphical models and sociological principles in three ways: modelling the environment, learning and predicting.

Concerning modelling, the work assumes that there are some parts of the observed scene which are more attractive to pedestrians and some areas, repulsive. By quantifying these "affordances" as a consequence of certain Points of Interest (POIs) and the different elements in the scene, it is possible to model this scene under observation with different costs as a basis of the features contained within.

Concerning learning, this work primarily extends the Growing Hidden Markov Model (GHMM) method - a variant of the Hidden Markov Model (HMM) probabilistic model- with the application of Prior Knowledge to initialise a topology able to infer accurately on "typical motions" in the scene. Also, the model that is generated behaves as a Self-Organising map, incrementally learning non-typical pedestrian behaviour and encoding this within the topology while updating the parameters of the underlying HMM.

On prediction, this work carries out Bayesian inference on the generated model and can, as a result of Prior Knowledge, manage to perform better than the existing implementation of the GHMM method in predicting future pedestrian positions without the availability of training

trajectories, thereby allowing for its utilisation in an urban scene with only environmental data.

The contributions of this thesis are validated through experimental results on real data captured from an overhead camera overlooking a busy urban street, depicting a structured built environment and from the car's perspective in a parking lot, depicting a semi-structured environment and tested on typical and non-typical trajectories in each case.

To my Mom and Dad,
without whom, none of this would have been possible.

Acknowledgement

This thesis was a journey on which I embarked on in early 2016. Completing this long and arduous journey could have never been accomplished without the presence and support of many people. First and foremost, I would like to express my gratitude to my advisors Anne Spalanzani and Dominique Vaufreydaz for their support in making this thesis possible and for teaching me to do research.

Many thanks to the people involved within the VALET project, the French Ministry of Education and Research and the French National Research Agency (ANR-15-CE22-0013-02) for having funded this reseach over the past three and half years.

I would also like to thank my reviewers Luis Merino and Thierry Chateau for their invaluable and constructive feedback on making this thesis better.

I am indebted to Alessandro, Mario and Victor for their timely guidance and advice for research as well as life. Victor, without your motivating talks in those early days, this thesis may have never been finished. Mario, your outlook on life and your readiness to help was an immense source of support during the times I needed it.

A big thank you to all the friends and colleagues from CHROMA, Pervasive and other teams at Inria Grenoble for all the help, discussions, an laughs that made my time there enjoyable. The long coffee sessions will be missed. Thanks too to Niranjan, Anshul, Rabbia, Daan, Vlad and Ozgur for your company - you guys made my stay there memorable. A shout out to all the office mates I've had over the years who have had to put up with my sometimes idiosyncratic habits. Thank you Jerome, JAD and Thomas for your help in collecting the data used for this research.

Speaking of friends, there are two who deserve special mention: David, for all the advice and the *long* gym sessions and Mathieu, for all the great discussions, thank you for the laughs and adventures.

Many thanks to Iris, for being a constant source of support and encouragement during the toughest parts of the thesis. Thank you for being there.

Last but not the least, I would like to thank my parents for their unconditional support during the entire thesis. Special thanks to my mother, Swarna, for all her support and sacrifices in making me who I am today.

Contents

Acronyms

APF Artificial Potential Field. 64

AVs Autonomous Vehicles. 1, 5, 7, 52, 141

EM Expectation Maximisation. 94

GHMM Growing Hidden Markov Model. 10, 11, 39, 41, 49, 50, 96, 97, 101, 109, 110, 117, 129, 137

HMM Hidden Markov Model. 32, 39, 49, 94

IBW Incremental Baum-Welch Method. 48, 49, 94

ITM Instantaneous Topological Map. 89

IV Intelligent Vehicle. 17

JPD Joint Probability Distribution. 31–33, 48, 92, 96

LiDAR Light Detection and Ranging. xvi, 2, 5, 94, 122–124, 133, 138

PKF Prior Knowledge Framework. 121

POI Point of Interest. xv, 10, 55, 62, 75, 76, 78, 104, 107, 112, 117, 122, 124–126, 128, 133, 137, 139, 141

PTM Prior Topological Map. 86, 88, 89, 91, 96, 97

S-ITM Semantically-Aware ITM. 81, 85, 86, 89, 91, 92, 96, 97, 105, 117

SA Situational Awareness. xv, 3–5, 141

VRU Vulnerable Road Users. 17

List of Figures

List of Algorithms

1 Introduction

> *You can't do better design with a computer, but you can speed up your work enormously.*
>
> **— Wim Crouwel**
> (Graphic designer and typographer)

In recent years, technology has improved by leaps and bounds bringing with it increased processing power and powerful sensors packed into ever smaller packages. All of these advances have made Autonomous Vehicles (AVs) feasible with some of them commencing commercial operations soon. Yet, one of the main obstacles and, indeed, worries of the general populace and the scientific community alike is guaranteeing safety with respect to other road users, especially the most vulnerable ones like pedestrians and cyclists. This poses a critical challenge since human behaviour is dependent on many factors (e.g., perception, intention, social standing, cultural biases, etc.). Motion planning for autonomous vehicles in a chaotic environment like an urban area remains an NP-Hard problem (Reif et al., 1994) made more complicated with the large number of individualistic elements that can be encountered. Estimating pedestrian intentions accurately becomes paramount to the successful deployment of autonomous vehicles. In this thesis, we study the case of pedestrian behaviour in urban areas under the influence of a built, structured environment in estimating their positions over time.

We begin by presenting, in this chapter, a general context and our motivation for the presented work in the domain of AVs. Following this, we formalise the problem tackled in this thesis, the proposed approach and the associated challenges. Then, in Section 1.4 we discuss the contributions of this work, finally presenting the outline of this manuscript in Section 1.5.

1.1 General Context

Traffic related fatalities are a global pandemic. By 2016, in absolute numbers, there were around 1.35 million deaths on the world's roads. Of this, vulnerable road users represent over half of these deaths with a disproportionate number of pedestrians and cyclists being affected while car occupants make up 29% of these deaths as presented in the World Health

Organisation (2018) report on global traffic accidents. This report goes on to provide insights and recommendations on the precautions and safety standards that could be employed by member countries of the United Nations Organisation. One of these is the ensure safe vehicles on the road.

Based on customer demand and strengthening regulations on this front, automakers have begun innovating in vehicular safety. Currently, there are many technologies on the market that provide *Automotive night vision*, *Emergency driver assist*, *Lane departure warning systems* and others that aid in keeping both the driver and the vulnerable road user safe while driving. These technologies, collectively called Advanced Driver Assistance Systems (ADAS) have been successfully used for the past few years to reduce mishaps on the road by sometimes taking control from the driver when collisions are imminent. It is under the ambit of these technologies that self driving cars have made their debut.

Autonomous vehicles hold the promise of a safer, tension-free commute for everyone involved. The Society of Automotive Engineers (SAE) has proposed a five level roadmap for the automation of cars on our roads (SAE International, 2018). The current revision of this proposed standard in autonomy can be summarised by Fig. 1.1. This infographic describes the technologies that define each level. Levels 0 to 2 require constant human attention and existing ADAS technologies suffice to meet the conditions prescribed to be classified under these levels. Levels 3 and above, on the other hand, require more.

The current state of technology on research platforms with their massive array of sensors - LiDAR, radars and cameras in different positions - has enabled us to reach an automation level of 3 according to the infographic. An interesting point to note here is that the standard prescribes that when the autonomous platform fails to perform, the *human* driver must take action. The implication here is that in many existing conditions, humans perform better than existing technologies. Given that these levels are sequential, we can assume that Level 4 should perform as well as humans in most conditions and Level 5 should necessarily surpass this performance under all conditions everywhere.

Very recent demonstrations by Tesla, Inc (2019) have showcased a possible move from the existing Level 3 capabilities to a commercially available Level 4 autonomous system. Yet, one of the criticisms of this demonstration was that it did not deal with urban areas or in chaotic environments.

A main reason for Level 4 and subsequently, Level 5 technologies being developed so gradually is the existence, or lack thereof, of regulations. A secondary, and more important to our discussion, reason is technical. The clear implication of Fig. 1.1 is the obvious influence of human drivers in the driving process. A move from Level 3 to 4 and ultimately to 5 must necessarily accommodate the interests and behaviours of vulnerable users on urban roads just as much as a human driver.

Consider, for example, the unfortunate incident in 2018 involving Uber's autonomous car and the fatality of a pedestrian. Reports of this incident have underlined the factors

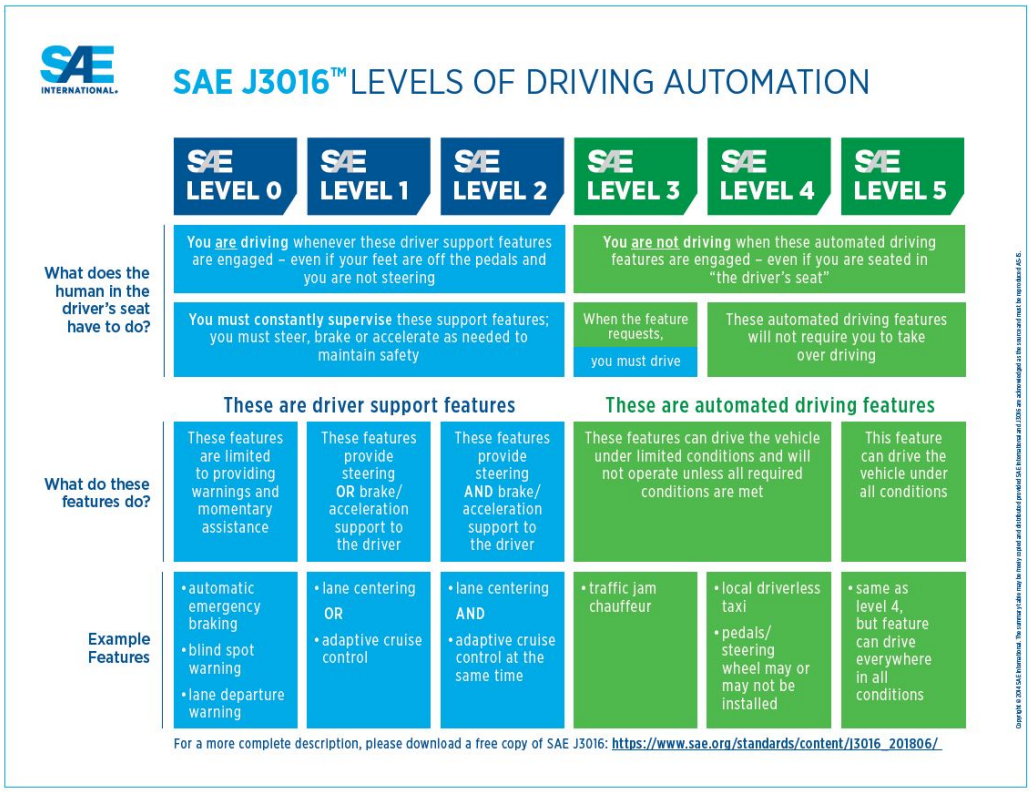


Fig. 1.1.: Levels of Autonomy as defined by SAE [Image credit: (SAE International, 2018)]

that caused the very first autonomous vehicle related death (Wakabayashi, 2018). Amongst other factors, the cause of this accident has been a cascading failure of *Situational Awareness (SA)*.

Situational Awareness as a concept was first presented in Endsley (1995) which described the different mental steps required by any intelligent agent to perform a task. This model describing Situational Awareness can be seen in Fig. 1.2. As we can see from this figure, SA is a combination of three levels following which a decision is taken and an action is executed leading to a new situational awareness, thus closing the loop for any task undertaken by an intelligent agent.

Consider once again the Uber incident of 2018. A woman walked perpendicular to the direction of motion of the autonomous car, crossing the street illegally while the environment was dark. This incident is illustrative of scenarios in urban areas involving vulnerable users. In this specific case, to attain the first level of SA, the car should have *perceived* the woman even while dark. The context of the situation, that there could be a pedestrian in the area should have been *comprehended* and, once an object had been detected, *predictions* should have been made about the future states of this obstacle. Sadly, the person was misperceived leading to the cascading failure of the entire system (National Transportation Safety Board, 2018). Had the system been operating at a higher level of Situational Awareness, this un-

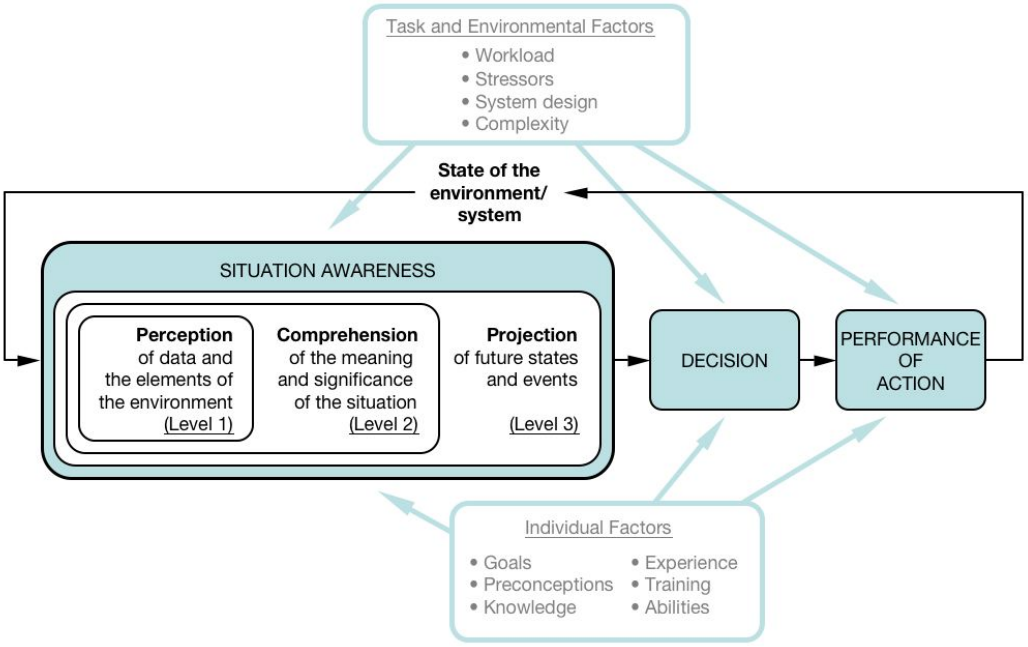


Fig. 1.2.: A model of SA, based on work presented in Endsley (1995). [Image credit: (Lankton, 2007)]

fortunate incident might have been avoided. Focusing on Figures 1.1 and 1.2 together, it becomes clear that a higher level of autonomy directly implies the capacity to infer at a correspondingly higher level of SA. Thus it becomes imperative to pursue human cognitive cues in designing systems requiring such Situational Awareness.

McKnight and Adams (1970a) and McKnight and Adams (1970b) devolve the task of urban driving into more elementary cognitive tasks. Within this ambit, a driver sequentially: looks at the scene that he is driving in, identifies the different features of this scene, decides the areas of the scene that is possible to drive in, identifies pedestrian and non-pedestrian lane markings and executes actions differently in residential and commercial areas. In commercial areas, he/she identifies areas of possible pedestrian density (*e.g.,* sidewalks, crosswalks, open areas where pushcarts can appear, building openings etc.) depending on the time of day (McKnight and Adams, 1970a). With a higher level of caution, the driver anticipates pedestrian behaviour in other areas such as drinking areas; sections of the street where the driver has the legal right of way but not the de-facto right; gives special attention to obscured areas and possible points of origin on the street. He/she then drives evenly through the area without disturbing other users of the street, vulnerable or non vulnerable (*e.g., swerving*) (McKnight and Hundt, 1971).This anticipation on the part of the driver of pedestrian intentions is possible because of a human cognitive capacity of "putting ourselves in their shoes." The driver identifies areas of pedestrian distribution in an area, notices a pedestrian and asks himself where the destination of that pedestrian could be given

the knowledge he possess of the area around him, allowing him to predict and take evasive actions or navigate around these dynamic obstacles if necessary.

The road to full automation is paved through increasing Situational Awareness of autonomous vehicles. No commercial deployment of AVs can ignore the challenges posed in operating in an urban area. As discussed earlier, AVs should perform as well as a human at the very least for any viable success. This implies a capacity of, not only reacting to vulnerable users, but also to anticipate them utilising their perspectives on planning and execution of tasks and negotiating a safe and successful passage amongst them. In other words, the vehicle must look to the human mentality of "putting ourselves in others shoes" to raise their levels of Situational Awareness. To arrive there we can identify the following main challenges:

1. **Reliable Perception** The first level of Situational Awareness (Fig. 1.2) depends heavily on robust and reliable perception. Current state of the art on autonomous navigation and perception prescribes heavy usage of powerful technologies such as LiDARs, cameras, Radar, etc. Some are complementary - where only one of these systems is insufficient to perform at required capacities. For example, Zhang et al. (2014) depend on the fusion of LiDAR data with that of a camera to detect vehicles in the scene. Such an approach, while sufficient under good weather conditions cannot compete under sub-optimal conditions such as cloudy skies, rain or snow. On the other hand, reliable semantic identification is becoming possible in images under many conditions using approaches such as He et al. (2017) and Redmon and Farhadi (2018). Drawbacks with these approaches are that a choice must be made between accuracy and speed since they are seldom real-time in identification. A second requirement for reliable perception is the need for accurate maps. Such maps play a pivotal role in localising the AV, perhaps using technologies such as GPS signals or through other means. Likewise, these maps also provide vital information on the area surrounding the AVs in creating a mental model which could be used as a fallback when the external perception systems fail, such as providing positions of commercial establishments, geometrical data on the scene, *etc.* Likely, future work in the field will focus on generating and maintaining good maps of the areas the vehicle is deployed in.

2. **Anticipating Pedestrians and assessing risk** The final level of SA is the prediction of future states of all the elements in the observed scene. These include dynamic obstacles such as cars, pedestrians, bikers and motorcyclists whose behaviours differ significantly within the same setting and certainly in different settings. For example, cars on a highway behave differently than in an urban area. The same is applicable for pedestrians and bikers in built and open environments. Thus, the autonomous vehicle must be able to anticipate the behaviour of these shared space users and predict their

positions in the future. In doing so, it must also estimate the level of risk to itself and to others in its vicinity.

This thesis addresses the second challenge, *i.e.*, that of anticipating and predicting pedestrian behaviour, specifically in built, urban areas.

1.2 Problem Description

Consider the motion of an AV in an urban area. The goal of this thesis is to develop a framework, from the perspective of an AV, to:

1. Model an environment conforming to pedestrian behaviour, without the observation of their trajectories
2. Capture their trajectories in new and previously unseen environments
3. Execute a learn-and-predict system where, when there are instances of non typical trajectories, such behaviours are learnt
4. Predict future pedestrian positions following typical and non typical motion patterns.

The Look-Once Principle As a contrast, we consider once more the behaviour of human drivers in urban areas. We have already discussed the cognitive aspects of driving in such situations in the previous section, specifically the ability of a driver to look at an environment and, intuitively, mentally model it under different circumstances. Within this mental model, drivers can infer pedestrian behaviour by "putting themselves in others' shoes." This extraordinary capability of humans in predicting others' position is a sociological phenomenon based on Schema (Casson, 1983). Humans look only once at a scene and glean information from it such as the different objects within and the interactions between them. Secondly, the human processing system is quick and accurate, allowing for complex actions to be executed within limited timeframes. These two ideas form the basis of the "Look-Once Principle" wherein a system is only compliant to this principle if it :

1. Requires no more than one iteration of a process to glean a maximum amount of information from the environment.
2. Optimises the execution of the process in such a way as to infer accurately the interaction of objects in the environment with little to no extra information other than the one provided by the first iteration.

In our proposed approach, we endeavour to follow closely the "Look-Once Principle" in modelling an predicting pedestrian behaviour in urban areas. Significantly, there exist differences in pedestrian behaviour in structured and unstructured urban environments.

Within this, pedestrian motion can be broadly classified into "Legal" crossings (or "Typical" motion) and "Illegal" crossings (or "Atypical" motion). In a structured environment, where each object within has a well defined function and convention, any one object behaving *un*conventionally, *for e.g.,* crossing on the street and not on a designated cross-walk, is considered "Illegal" behaviour while the contrary is considered "Legal".

Our approach to this problem is to build prior knowledge, for a system with regard to the environment it is operating in based on the different features within which can impact pedestrian behaviour. Following this prior knowledge, the system can then update this knowledge with new information as it comes along. In possessing this knowledge, either prior or posterior, the system should be able to answer the question "With a pedestrian identified in the observation space, knowing their current trajectory, is it possible to predict their position with a long term horizon and their possible goal?"

1.2.1 Challenges

The objective of this thesis, expressed as three problems – that of modelling the environment, learning pedestrian behaviour to generate Prior Knowledge and leverage it to predict motion in a human-like manner – comes with its own set of challenges in approaching these problems.

Unreliable Observations: Prediction of pedestrians in any environment requires reliability in the tracking of said pedestrians. Yet, sensors operating in noisy environments capture unreliable observations. The huge problem in this domain is the data association problem, whereby a tracker, on encountering an occlusion, can lose track of the object under observation. Once the pedestrian is re-observed, they may not be assigned the same tracking ID. In most existing approaches dealing with prediction, the ground truth of the trajectory is assumed to be known from start to destination (Vasquez et al., 2009; Ziebart, Ratliff, et al., 2009; Kitani et al., 2012).

Learning Constraints: A learn-and-predict system where trajectories train a model all the while its previous iteration is used to predict another trajectory's future (Vasquez et al., 2009) require a large number of training trajectory observations to reliably used for prediction. Thus, such an implementation cannot be used, for example, on moving platforms like AVs where, on deployment, the learn-and-predict system cannot observe sufficient trajectories while moving to perform online learning and inference. Another drawback would be the loss of information on the trained model once the AV moves from that specific area.

Transfer of Learning This is a challenge affecting the problem of modelling the environment as well as pedestrian prediction. Consequent to the previous challenge mentioned, the

parameters and the model learnt are wasted when the platform localises itself in a novel view. Under such scenarios, the model is required to retrain itself by observing new trajectories over time which will once again be lost when the platform moves. Some approaches in literature have attempted to evade this recurring issue by attempting to transfer the learnt model to novel scenes (Kitani et al., 2012; Ballan et al., 2016). These approaches discretise the observed environment into constituent, recurring features over different scenes (*e.g.,* road, pavement, *etc*). These features are associated with a weight that is learnt on observing behaviours which are then generalised and applied in other scenes. However, such knowledge transfer only becomes straight forward when there are large datasets available (C. Liu et al., 2009).

Dearth of Training Datasets The previous challenge leads directly to this one wherein the transfer of learning is affected by the dearth of good training datasets in urban areas depicting pedestrian prediction (Ridel et al., 2018). This challenge can be solved by two methods – creating new datasets and attempting to utilise online learning mechanisms. In considering the former, the learnt models are rooted to specific scenarios and behaviours leading to other challenges for generalising this learnt model. The latter is almost always computationally expensive.

1.3 Proposed Approach

Our proposal in solving the challenges presented in the previous section can be found in Fig. 1.3. This approach consists of various cascaded modules described below.

Sensor data coming from the environment are parsed through a "Perception Layer" which performs the task of identifying and segmenting the different features of the environment. Parallelly, a tracking module identifies moving objects in the scene, tracking them over time in both position and velocity. The outputs from these two modules then feed into the other modules sequentially.

1. **Environmental Modelling:** The aim of this module is to accrue information about the environment where the observer is present and to convert this semantic information into quantifiable values. This is the first step in building Prior Knowledge.

2. **Behaviour Learning:** This module converts the quantifiable values into graphical models that can be used to represent typical and atypical trajectories of pedestrians. Internally, this module utilises the tracker information to update the representation of the graphical model to more accurately model pedestrian behaviour. The resulting graphical model contains embedded within it information on the environment as well as the probabilistic distributions of pedestrian motions at each of the states defined by the topology.

3. **Pedestrian Trajectory Prediction:** Within the context of our defined problem, prediction of pedestrian motion happens over a specified horizon. Performing exact Bayesian inference over the graphical model results in a probabilistic distribution over the observed area for each pedestrian. In the broader context of Situational Awareness, this probability distribution aids the decision making process, which is beyond the scope of this work.

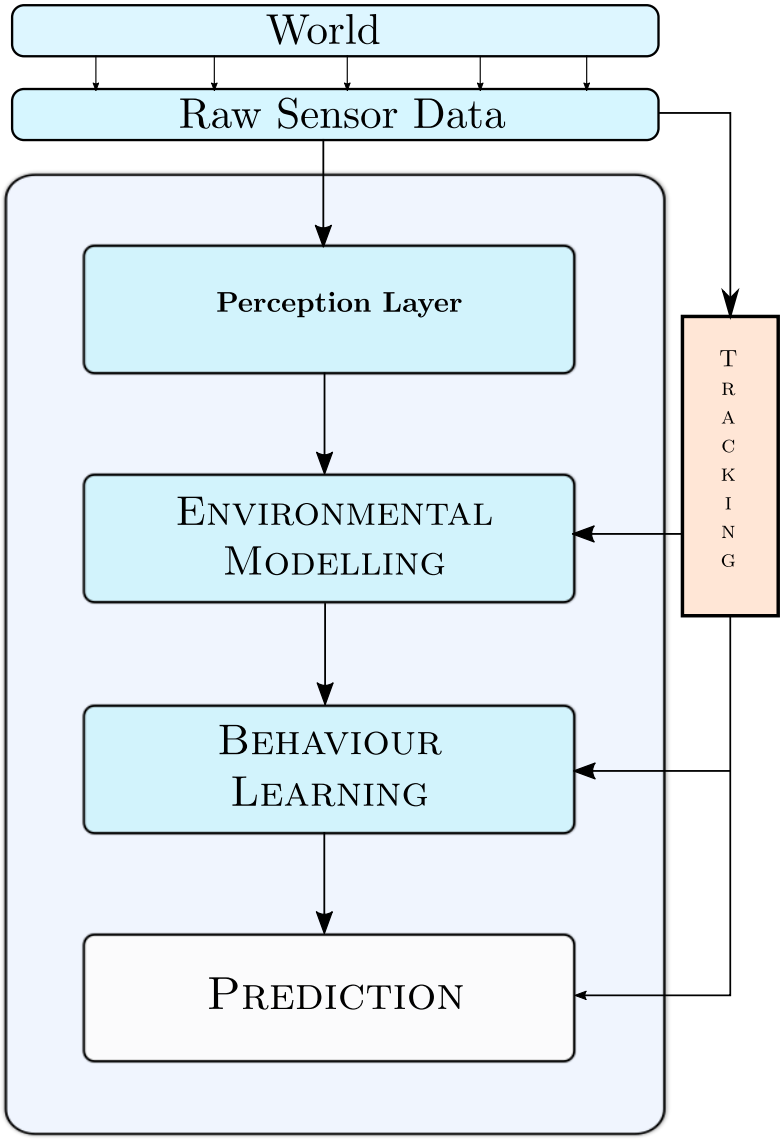


Fig. 1.3.: Proposed approach to solve the pedestrian prediction problem

1.4 Contributions

The main contribution of this thesis is a human-like prediction framework, with the capacity to learn, for predicting pedestrian behaviour which can be deployed in an autonomous vehicle operating in built, urban areas.

This main contribution can be divided into three sub-contributions as follows:

1. **An algorithm to model the environment of a built, urban area composed of different features as a cost function:** The proposed algorithm leverages the sociological principles of *Natural Vision* and *Natural Movement* (Gibson, 1979) to model the different features present in the environment and assign costs from a pedestrian's perspective to various areas of the scene. This approach differs from existing Inverse Reinforcement Learning (IRL) approaches (Ziebart, Maas, et al., 2008) for solving the same problem in the manner of defining the weights of the observed and identified features. While IRL techniques depend on previously observed pedestrian trajectories to learn these weights, the proposed method utilises sociological cues in assigning weights to perform at the same measure. Furthermore, this approach employs the abstraction of the attractiveness and repulsion of different areas via the potential fields to define the effects of the identified features on the behaviour of a pedestrian in the built, urban area. It also introduces the idea of attractive Point of Interests in an observed area as the destinations of pedestrians in the environment, thereby leading to a cost function independent of prior observations of pedestrian behaviours in the said environment.

2. **Extensions to the Growing Hidden Markov Model (GHMM) approach utilising prior knowledge aimed at its application in an autonomous vehicle.** In this thesis, we propose two extensions to the GHMM (Vasquez et al., 2009) method to apply this method in new and previously unseen areas with none to few observations to train the underlying graphical model. These extensions adapt the topological and structural phases of the GHMM respectively. They are as follow:

    a) **An extension to the topological mapping of the environment.** This thesis proposes an algorithm to generate a topological map corresponding to the contours of the observed environment. This topological map exploits a generated cost map to *a priori* capture typical trajectories in this given environment, before the observation of any pedestrian trajectories. This topology can, however, change during the course of the lifetime learning process of the GHMM. The initialisation of this prior topological map aids in the explanation and prediction of non-typical trajectories which would have been missed in the existing implementations of this method.

b) **An extension to the structural learning of the underlying graphical model.** The proposed extension modifies the process in which the structure and parameters of the underlying graphical model are learnt. It manipulates the states and their interactions bringing them in line with the topology of the environment by using the previously developed Prior Knowledge from the aforementioned two contributions. Furthermore, this extension drastically reduces the training data required for accurate predictions of pedestrian trajectories by initialising the graphical model with realistic priors, which is an advantage over the existing implementation.

3. **An Extension to the Growing Hidden Markov Model (GHMM) approach to infer pedestrian positions in the presence of dynamic obstacles.** Finally, within the context of our problem, we propose a method to integrate a variety of moving objects in the environment into the prediction process of the GHMM method. This method utilises the generated environmental model to confine a pedestrian's trajectory to predict their future positions based on the current state of the environment.

1.4.1 Papers Published

During the course of the development of this thesis, the following scientific papers were produced.

1. Vasishta P., Vaufreydaz D., Spalanzani A., "Natural Vision Based Method for Predicting Pedestrian Behaviour in Urban Environments," In *Proceedings of the IEEE 20th International Conference on Intelligent Transportation Systems (ITSC2017)*, Yokohama, Japan, 2017.

2. Vasishta P., Vaufreydaz D., Spalanzani A., "Building Prior Knowledge: A Markov Based Pedestrian Prediction Model Using Urban Environmental Data," In *Proceedings of the IEEE 15th International Conference on Control, Automation, Robotics and Vision (ICARCV2018)*, Singapore, 2018. ***(Won the Best Student Paper award)***

3. Vasishta P., Vaufreydaz D., Spalanzani A., "Urban Pedestrian Behaviour Modelling using Natural Vision and Potential Fields," In *Proceedings of the 9th Workshop on Planning, Perception and Navigation for Intelligent Vehicles at the IEEE International Conference on Intelligent Robots and Systems(PPNIV-IROS2017)*, Vancouver, Canada, 2017.

1.5 Thesis Structure

Part One: Background and State of the Art

Chapter 2: Intentional Motion Prediction This chapter presents a review of the state of the art in intentional motion prediction with an emphasis on modelling pedestrian behaviour and predicting their trajectories under uncertain environments.

Chapter 3: Probabilistic Models This chapter deals with presenting the basic mathematical theory, notations and concepts of probabilities, probabilistic states and graphical models and a quick primer on Hidden Markov Models.

Chapter 4: Growing Hidden Markov Models In this chapter, an introduction to the Growing Hidden Markov Model tool is presented, an extension of which, will be presented as one of the proposed contributions of this thesis. This chapter discusses the basic methodology used for lifelong learning and subsequently the drawbacks inherent to this method.

Part Two: Modelling and Prediction in Human Populated Urban Areas

Chapter 5: Modelling the Human Environment This chapter commences with a discussion on the sociological influences on pedestrian behaviour. An algorithm that maps the observed environment to potential costs for a pedestrian in an urban area is proposed. A validation on a typical urban scene is also presented.

Chapter 6: Learning and Prediction of Motion Once a cost map for an observed environment is available, it can be used to explain typical trajectories in the environment. As soon as there are new observations available, this approximation of typical trajectories can be updated and new trajectories learnt to perform better inference. This chapter presents an algorithm that utilises information about the environment to generate a dynamic graphical model that is used to perform inference and predict future pedestrian positions, including prediction under dynamic conditions.

Part Three: Implementation and Experiments

Chapter 7: Experimental Results - Exoperception This chapter first presents the architecture of the proposed "Prior Knowledge Framework", describing the pipeline that combines the approaches of generating a cost map and the utilisation of the graphical model corresponding to this cost map. It also presents a dataset, wherein the data is captured from an external, static observer, which is then used to perform the different experiments. Qualitative and quantitative results are presented for different scenarios of pedestrian behaviour as

well as for those scenarios that include dynamic obstacles in the form of other shared space users.

Chapter 8: Experimental Results - Egoperception In this chapter, we present results on the efficacy of our proposed framework from sensors located on an autonomous vehicle placed in a semi-structured area. First, the dataset and the platform used to capture this dataset is presented. Qualitative and quantitative results obtained from the application of the Prior Knowledge framework on this dataset are presented.

Part Four: Conclusion

Chapter 9: Conclusions and Perspectives Finally, this chapter presents some concluding perspectives on the work, a summary of the contributions with potential ideas for future studies on the topic.

Part I

Background and State of the Art

Intentional Motion Prediction

2

> *If I have seen further it is by standing on the shoulders of Giants.*
>
> **— Isaac Newton**

In this chapter, we review the state of the art in increasing the Situational Awareness from the perspective of an autonomous vehicle plying on urban roads. In Chapter 1, we introduced the problems and their contexts in view of this thesis. In increasing its own situational awareness, an IV should perform the tasks modelling the environment, learning the behaviours of other shared-space users and then predict their intentions. This problem, also called the motion prediction problem, involves inferring also the intentions and interactions between these shared space users over future horizons. These shared space users can broadly be classified as Vulnerable Road Users (VRU) and non-VRUs. By definition[1], VRUs are "non-motorised road users, such as pedestrians and cyclists as well as motorcyclists and persons with disabilities or reduced mobility and orientation." Predicting the behaviour of non-VRUs and their interactions have been surveyed by Lefèvre et al. (2014) and are beyond the scope of this work and thesis. Instead, we limit ourselves to a discussion on the prediction of VRUs, specifically the prediction of pedestrian motion and behaviours in urban areas.

Classifications of existing methods vary from study to study (Ridel et al., 2018; Brouwer et al., 2016; Rudenko et al., 2019) but they can be broadly categorised as: physics-based methods, pattern-based methods, contextual-cue based and interaction-aware methods with each of these categories being treated in the following sections.

A note on semantics

For the rest of this chapter and the thesis, we will use the semantics of pedestrian behaviour as set out by Vasquez (2007). We follow the convention of scientific literature, as mentioned in there, where the terms "behaviour" and "motion pattern" refer to different ideas - behaviour is intentional motion towards a destination while pattern is repeated observation of the same behaviour. These destinations are states – either observable, as in the case of

[1] https://ec.europa.eu/transport/themes/its/road/action_plan/its_and_vulnerable_road_users_en

a specific point in a spatial map or, inferable at a higher level, such as “idle walk”, “brisk walk”, “stop”, “wandering” (Kanda et al., 2009).

We also make no distinction between the terms pedestrian motion and pedestrian behaviour. This seeming inconsistency can be explained away by sociological assumptions that “all human motion is directed” which McFarland (1989) defines as “goal-directed” behaviour. Thus, if all motion is directed towards a goal, according to our chosen convention, this becomes behaviour. Hence, these terms will be used interchangeably in this thesis.

2.1 Physics-Based Approaches

Physics-based approaches incorporate some of the simplest mechanisms for predicting motion in an environment. In most of these approaches, prediction is performed by the application of explicit functions that model future positions, velocities and angles based on current observations. While these approaches are used for prediction, their origins more often than not come from applications developed for tracking objects. Typical models among in such literature include kinematic models utilising Constant Velocity (CV) models, Constant Acceleration (CA) or Constant Turning (CT) models for prediction purposes. P. Blom (1984) developed a method for tracking moving objects, called the Interacting Multiple Model (IMM) method, whose modes undergo abrupt changes between the above mentioned typical models. An early effort to extend this filter and capture the cross and along-acceleration of objects being tracked was proposed by Best and Norton (1997) being applied to curvilinear motion of these objects. Another approach (A. Elnagar, 2001) utilises a Kalman filter with an assumption that the object moves with a constant velocity. Thus a pattern begins to emerge. These approaches do not tackle the problem of learning, resorting to handcrafting the values of the various parameters if required. Furthermore, they limit themselves to simplistic prediction problems of short horizons with an expectation of complete knowledge of the world.

A shift from the *sense-predict* cycle to *sense-learn-predict* cycle can be noticed in the work of Ashraf Elnagar and Gupta (1998) where it is expected that knowledge of the current and previous positions of all the objects in a 3D dynamic environment are available. This knowledge is then utilised to optimise parameters of a third order Auto Regressive Model (ARM) describing moving object trajectories for short term prediction of the objects in the scene. A similar method approach is followed by Cai et al. (2006) for tracking and predicting multiple objects in a scene, A second order ARM is used to estimate trajectory parameters which are then fed to a particle filter to arrive at short term prediction of objects being tracked. While the former initiates a study into prediction under dynamic environments and the latter extends it to prediction in the multi-object domain, both suffer from a similar limitation. Neither of these approaches accounts for changes in object trajectories in the scene. An exploration in this direction, albeit for vehicles, was proposed by Kaempchen

et al. (2004) to predict stop and go situations for the objects being tracked. These situations arise frequently while driving on roads as a response to external stimuli – braking of the vehicle in front, crossing pedestrians, stop signs, traffic lights, *etc.* – which are modelled via the aforementioned IMMs between constant velocities and constant accelerations. Similar motions are observed amongst pedestrians, if perhaps, not for the same factors thus driving research in the direction of including environmental factors. Subsequently, Pellegrini et al. (2009) proposes a Linear Trajectory Avoidance (LTA) model for short term prediction of pedestrian trajectories in which the intentions of pedestrians are explicitly included. Also inherent in the model are interactions with other moving objects in the environment and their effects on the trajectory being predicted. Pedestrian trajectories are modelled as energy functions directed at an intention (here, a destination which could exist outside the field of view) on which their movements are dependent. The modelled function's parameters are then learnt based on observed trajectories whereby these parameters are optimised using Gradient Descent. Prediction is performed using this trained model. This work is in close association with the Social Force model expounded by Helbing and Molnar (1995) in which every pedestrian has reactive social space that affects their motion. While Pellegrini et al. (2009) worked on prediction of targets individually, Yamaguchi et al. (2011) extended this concept to include the behaviours of groups. More importantly, it also included the effects of the social force model as an inherent part of the energy function describing the motion of these groups. The energy function developed in this work emphasises on the knowledge of hidden states of a personal nature such as a preferred speeds, destinations within the scene and social groupings. The parameters associated with these states are learnt from observed trajectories in a scene using Support Vector Machines (SVMs). This model, however, does not ascribe to the tracked pedestrians more than one behavioural state. For example, irregular motions such as start and stop behaviours, which are common pedestrian behaviours, are not captured. In Keller and D. M. Gavrila (2014), a model is proposed to tackle this limitation. It proposes a Gaussian Process Hierarchical Model (GPDM) trained separately for walking and stopping motions on a variety of trajectories. A hierarchical framework is presented which utilises the identification of the behaviours via the GPDM to statistically match trajectories, thereby predicting pedestrian positions. In the same spirit of this work, Kooij et al. (2014) proposed a Switching Linear Dynamic System (SLDS) to predict pedestrian trajectories from the perspective of an autonomous vehicle. The utilisation of this model was able to capture more accurately unusual behaviours of pedestrians. An SLDS allows for the dynamics of the trajectory to change at every time step. Learning is performed via Bayesian inference of the posterior distribution. Other approaches (Batkovic et al., 2018) introduce environmental constraints derived from a local map to estimate pedestrian positions using simple dynamics.

From the works discussed in this section, an evolution of models can be seen from simple applications of physics based formulations to higher levels of inference of behaviours

such as stop and go to the introduction of maps to the prediction process. These physics based approaches are utilised because of the relative ease and simplicity of their implementation. Better accuracy in prediction requires introducing more complex models leading to a requirement of learning the parameters that establish this complexity.

2.2 Pattern-Based Approaches

Pattern based approaches follow the sense-learn-predict pipeline to discover statistical behaviours among trajectories. These approaches, in contrast to physics based approaches, are heavily data dependent, requiring many instances of trajectories to learn the intricacies of human motion.

Early approaches into this domain involved clustering algorithms such as the one proposed in Kaufman and Rousseeuw (2009). Observed trajectories would be clustered and trajectory prototypes would be availed. Applications of these simple clustering algorithms had limitations of being unable to anticipate anomalous trajectories if such an exemplar did not already exist. A similar approach was proposed by Hu et al. (2006) where a *K-means* algorithm is used to cluster trajectories. While this work was applied to the behavioural prediction of cars in urban areas, the algorithm developed can be used for pedestrians. Here, training trajectories were clustered in a manner to provide hierarchical piecewise trajectories over the observed area. A moving object follows these piecewise trajectories and can easily switch from one behaviour to another at the end of every piecewise trajectory. Such an approach captures common behaviours such as turning or continuing along the same path at an intersection. Prediction is performed by estimating the probability of choosing the piecewise trajectories by the moving object. Then, the position of this moving object over a time horizon is a dynamics based extrapolation of object on this partial trajectory. Consequently, anomalous behaviour in the scene can be detected when the tracked object performs motion that does not correspond to motion along the learnt piecewise trajectories. While this approach operated directly on trajectories to find abnormal behaviour, Varadarajan and Odobez (2009) proposed a model towards the same goal but operated not on the trajectories but on the semantics of the observed area. Here, they propose a probabilistic Latent Semantic Analysis model that learns from observed trajectories in an area, the semantics of this area. The vocabulary of these semantics represent segments of the area where pedestrians and cars move or perform stop & go motions. This model also learns segments where velocities of the moving objects are slow or fast. With these segments learnt, they utilise as a measure the Bhattacharya distance to estimate how far the tracked behaviour from what is considered normal for the segmented area in which this object is operating. While these approaches can be used to determine whether or not a moving object executes an anomalous behaviour, they fall short in performing long term prediction.

These approaches also do not take into account the intentions of the moving objects in the observed scene.

The path between two points in any environment can be established by more than one trajectory. Common clustering algorithms cannot perform this discrimination and instead result in a "mean" trajectory from amongst all the training trajectory data even if cases of trajectories exist in the training set that follow multiple paths to get to the same destination. One approach that works towards rectifying this oversight is presented in Ellis et al. (2009) which uses Gaussian Process Regression to learn pedestrian tracks for long term prediction of trajectories in static environments. In this approach, a motion model is generated for pixel-wise instantaneous velocity across the entire observation space which allows for probabilistic inference over a medium to long time horizon. A similar approach was taken in Joseph et al. (2011) where a Dirichlet Process was used to generate the priors for a Gaussian Process (DPGP) to generate motion patterns for a target object. Solving problems with DPGPs, however, become intractable quickly due to their high complexity thus limiting their usage when the area under observation is large. In response, Chen et al. (2016) presented their method, called the the augmented semi-nonnegative sparse coding (ASNSC) method to reduce the computational time of prediction while not sacrificing accuracy. In this method, the observed environment is discretised and trajectories are treated piecewise to arrive at its vectorised representation for each cell of the environment. On training this method with many trajectories, local motion patterns are discovered. The transitions between these local motion patterns are learned via a GP. In a step reminiscent of Markov chains, on the reception of an observation, the trajectory is back back propagated to assess which motion pattern generated it. This is then propagated forward in time to achieve prediction goals. Habibi et al. (2018) extend this method by introducing environmental contexts into the ASNSC process. The introduction of this semantic context allows the method to learn other parameters such as distance to crosswalk to generate a more accurate motion primitive for the discretised environment leading to a significant improvement in prediction accuracy.

Another method, proposed by Deo and Trivedi (2017) utilises a Variational Gaussian Mixture model (VGMM) for trajectory prediction from the perspective of a vehicle. This fully unsupervised method is able to discover goals in the observed scene as well as predict future positions of pedestrians. The starting and ending positions of all the training trajectories are clustered to find destinations in the scene resulting in subclasses. A separate VGMM for each destination is trained on the trajectories to ascertain common behaviours of each of these subclasses and is able to generate typical trajectories between each source and destination.

From the works presented in this section, is can be clearly seen that the introduction of context leads to an increase in the accuracy of modelling and prediction of pedestrian

trajectories. A point to be noted more is the need for many data points without which these statistical motion representation methods fail to accommodate atypical trajectories.

2.3 Context-Based Approaches

Context based approaches can be classified as those approaches that utilise social cues to perform long and short term prediction. These cues may also be corporeal cues such as head orientation, body pose *etc.,* that could point an observer to predict certain behaviours in pedestrians. Generally, these cues aid in detection of intentions, mostly in determining whether a pedestrian would preform a specific action or not.

One of the first approaches based on these cues was Köhler et al. (2012), which utilises human gait mechanics to predict pedestrian intention to step on to the street. The ultimate goal of this work is to identify early as to when a pedestrian standing at the kerb will initiate a movement on to the street. A sound sociological basis exists for this problem. Identifying biomechanical traits such as movement of the legs and forward bending of the torso can point to an initiation of motion. To identify these from a monocular video stream, this method proposed an HOG based identifier called the MCHOG long with an SVM to first identify the different parts of the body that move when initiating walk and the SVM to classify them as important or not. Another work in the same domain was presented by Goldhammer et al. (2013) to rectify some of the drawbacks of the previous method, Not only does it perform early identification of pedestrian intention, it also performs short-term prediction. To identify when a pedestrian initiated motion, the change in the velocity of his center of gravity is observed. On initiation of motion, the trajectories they will take is predicted based on a piecewise linear model. The presented methods utilised only the corporeal cues and did not consider any other factor that would affect this behaviour. Most importantly, these approaches do not consider the effect of the environment. For example, there is a higher chance of a pedestrian executing a crossing if he/she is moving in the direction of a crosswalk. Bonnin et al. (2014) provides a model to tackle this specific problem. This work proposes a hierarchical context tree to warn drivers of pedestrian intentions in inner-city areas and specifically at crosswalks. A list of 12 features relative to the car being driven is established and a single layer perceptron is trained. On testing for intent, the context tree decides on whether to or not to activate the crosswalk model on the detection of a crosswalk. If no crosswalk is detected, then the general inner-ciry model remains active to predict a crossing.

While these approaches work well for pedestrians already close to the kerb or near a crosswalk, consider someone walking along the sidewalk, some distance away from the crosswalk. This person could initiate a crossing either legally or illegally based on different cues exhibited before approaching the kerb. The discussed methods do not consider the history of motion of the pedestrian or previous cues. To fix this oversight, Andreas Th

Schulz and Stiefelhagen (2015) proposed a method involving Latent Dynamic Conditional Random Fields (LDCRF) which can work with trajectories of arbitrary time lengths. The model is trained in two parts – the first part trains on the dynamics of the trajectories such as the velocities and the second part specifically trains on the head pose of the pedestrian over the lifetime of the trajectory. It also trains to capture the time to event, where the event is the crossing action at the crosswalk. A correlation is made between the dynamics, the head pose and the time to cross by the proposed model which is able to discriminate a variety of behaviours based on this correlation. The authors envision the integration of this intention recognition model with one specifically designed to predict pedestrian motion as proposed in Andreas T Schulz and Stiefelhagen (2015). This integrated model, utilising IMMs for trajectory prediction and LDRCFs for intention recognition, performs better at predicting behaviours such as "stopping" and "crossing" leading to better position predictions. Quintero et al. (2017) performs a similar function, but utilises other corporeal cues such as shoulders and other skeletal joints. These joints are then trained via an HMM to recognise motions that initiate a crossing, while the same class of joints are passed through a Convolutional Neural Network architecture in Ghori et al. (2018) to achieve the same goal of predicting crossing intention.

2.4 Interaction-Aware Approaches

Interaction Aware approaches can be classified as a part of the *sense-understand-learn-predict* paradigm, whereby the model is not only cognisant of the environment it is operating in, it is also aware of the interactions the pedestrians and other moving objects have with each other and with the environment.

In literature, it can be observed that these approaches are broadly divided into two classes – those that use Markov Decision Processes (MDPs) and their derivatives and those that utilise Dynamic Bayesian Networks (DBNs).

MDP based approaches

Most MDP based approaches into pedestrian prediction are coupled with a planning phase where the motion of the observed agent, in this case the pedestrian, affects the decision that needs to be taken by the observer. One of the first approaches into this domain can be found in Bandyopadhyay et al. (2013). This work proposed an MOMDP (Mixed Observability Markov Decision Process), which a class of Partially Observable Markov Decision Processes (POMDPs), that learns to model pedestrian intention. It is assumed that this pedestrian has a finite number of intentions and that there are attractive goals in the scene that drive the pedestrian towards it. These intentions of the pedestrian are modelled as an MOMDP, trained offline and a policy for the observed environment arrived at. The action

space of the observer, in this case an autonomous golf cart, is limited to a few behaviours. In the online phase, when inference occurs, the golf cart reacts to the pedestrian's motion and takes a decision based on the inferred intention. However, the posterior projection of the intention is not connected to the semantics of the scene. It is assumed that the pedestrians move in an open world, in the direction of their goals.

Vasquez (2016) presents a method that utilises MDPs and combines it with semantic notions of the environment. Existing MDP based prediction and planning algorithms, designated MDP-Motion Prediction (MDPMP), are used to generate a costmap in an environment whose features are known. Their corresponding weights are also learnt via Inverse Reinforcement Learning (IRL). On this costmap, inference is performed via the proposed method. This method aims to predict accurately the goal as well as future positions over long horizons. The graphical model is devolved into velocity prediction and goal prediction. On superimposition, this leads to a more robust, long term trajectory planning. Another method proposed by Karasev et al. (2016) proceeds to integrate semantic maps into the inference process where an MDP's policy for the given environment is learnt corresponding to its semantics. The model has two parts – the planning part that works offline and a posterior estimation process that uses a Rao Blackwellized filter to perform the actual prediction part. Environmental cues are used to limit the prediction distribution spread. Similarly, better accuracy in prediction was achieved by also including the body orientation into the prediction process.

DBN Based approaches

Dynamic Bayesian Network approaches are frequently used in literature to model latent states in a time varying environment. Latent states, especially in the context of pedestrian trajectory prediction, take the form of intentions that cannot be directly observed. DBNs are also popular since they can also be used to model interactions between different dynamic objects in the scene. DBNs are also versatile since they can be thought of as generalised Bayesian Filters with HMMs and Kalman filters being its specific implementations. More importantly, many of the tools developed for learning and inference of these specific implementations can be reused for DBNs.

One approach, applied to predicting diver behaviour is found in Agamennoni et al. (2012). The agent's dynamics is conditioned on the context of the other drivers in the scene. This context is a hand crafted function used to capture the the interplay between the dynamics of the different actors in the scene. Such a function can try to fit distance from other actors, speed and direction of motion, intention *etc*. These weighted functions learn their weights based on observations using the Expectation Maximisation (EM) and inference is performed over this trained graphical model. While this model was used to perform long term prediction on cars under reduced or stressed sensory data, the same principle can be applied to

human actors in urban areas. Another approach, by Rehder and Kloeden (2015), subsumes this context in the form of a map and models a latent variable as the intended destination of the pedestrian. An occupancy map is used as a prior to represent the environment and the goals of the pedestrian are not known and are learnt online. Prediction is executed between the current position and the estimated goal for each position using the observed dynamics of the actors. A limitation of this method rests with the poor definition of the environment for performing inference. Occupancy grids offer no more information other than representing regions of the environment as free or occupied. Richer information could be used to model more complex DBNs. One such approach that takes into account a more detailed representation of the local environment can be seen in Hashimoto et al. (2015). This information pertains to the interaction between an intersection and a pedestrian waiting to cross. It also includes information on the traffic signals at this intersection. Specifically, this work proposes to model pedestrian intentions of slowing down and speeding up when close to the signalised intersection. The intended behaviour of the actor is inferred via particle filters.

Finally, there are other works that account for the interactions between the different actors in the environment based on Neural Network approaches. A seminal work in this regard is seen in Alahi et al. (2016). This work proposes an architecture based on a special case of Recurrent Neural Network (RNN) called the Long Short-Term Memory (LSTM) network. This class of architecture is able to retain memory of previous timesteps, rendering it appropriate for use in areas of sequence prediction such as that of pedestrians. There are hidden states present in this architecture that account for the "social" behaviour of people such as being influenced by other people or objects in the scene. This network is trained on position of tracked objects over time, learning "social behaviours" along with the semantics of its current position based on an occupancy map. This model is then used to infer future positions of trajectories. This work was extended by Su et al. (2017) to include crowds in the process. They achieve this by integrating Gaussian Processes to learn the correlation between social trajectories as learnt and inferred by the LSTM network.

Probabilistic Models

3

> *Users do not care about what is inside the box, as long as the box does what they need done.*
>
> **— Jef Raskin**
> about Human Computer Interfaces

In a chaotic world, there are very few times when one can take a decision with absolute certainty. In the field of robotics, especially in the domain of autonomous vehicles, there is always a hesitancy in proffering the absolute. Decisions are taken with varying levels of confidence in the observations as much as in one's own predictions about the state of the world. In the material to follow, probabilities - especially the Bayes' filter and its applications - will play a prominent role. Thus, in the interest of a self contained work, we begin with an introduction to the probabilities used in this work and a familiarisation with the notations used within. This chapter is influenced by the work of Vasquez (2007), upon which we improve. We also follow, in terms of notations, the work of Thrun et al. (2005) which the inquisitive reader can refer to for an in-depth treatment of the concepts discussed here.

3.1 Variables and States

A probability represents the belief in an event and can be posed as a proposition - " How likely is an event v to occur?", represented by $P(v)$. If $P(v) = 0$, the proposition v occurring is false. On the other hand, if $P(v) = 1$, then the proposition definitely occurs. Given these two extremes, the function $P(\cdot)$ maps this proposition such that $P(\cdot) \mapsto [0, 1]$.

The event v, called the variable, in a probabilistic environment, represents all the features of that environment. Imagining that all probabilities are answers to questions, one can think of variables as being the construction blocks of those questions. A discrete variable $\mathcal{V}$ can take on only the values present in the set $\{v_1, v_2, \ldots, v_n\}$.

Any given environment can be defined by its *state*. The state $\mathcal{X}$ is defined by a set of values it can take - $\mathcal{X} = \{x_1, \ldots, x_n\}$ where $x_1, x_2, \ldots, x_n$ are called the state variables. Some of these state variables can remain unchanged while others can vary with time. A *dynamic state* is one where the state is changing while a *static state* is an unchanging one.

We use the notation x_t to represent the value of the state X at time t. While referring to probabilities, by abuse of notation, we will omit the full definition of probability of state X at time t i.e., $\Pr([X = x_t])$ and denote it as $\Pr(x_t)$ or $P(x_t)$

Another way of distinguishing states is by defining the property of $completeness$. Consider a state x_t. This state is considered $complete$ if the information encapsulated within it can describe perfectly the entire environment. That is to say, by knowing only x_t, and disregarding all the information and the states that led to it, one can reasonably predict the future states $x_{t+1}, x_{t+2}, \ldots, x_N$. A complete state is only theoretically possible. The real world is inherently $stochastic$ and any attempt to capture the precise state of the environment is impossible due to the noisiness of any sensor used. Thus, in the definition of any state variable, there is a need to compensate for the noisiness of the sensors used. One can, however, approximate a complete state under specific conditions. With this approximation, a *Markov Chain* can be described - where the evolution of future states depend only on the current state and past states do not influence future states.

The basic laws regarding probability from which evolutionary laws are derived are presented here.

Independence Two random variables X and Y are considered independent ($X \perp Y$) if and only if

$$P(X, Y) = P(X)P(Y) \tag{3.1}$$

Conditional Probability The Conditional Probability $P(X \mid Y)$ i.e., the belief in X occurring knowing that Y occurs is defined by:

$$P(X \mid Y) = \frac{P(X, Y)}{P(Y)} \tag{3.2}$$

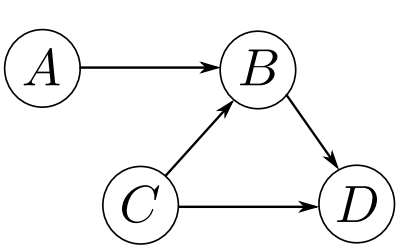
Equation (3.2) holds true only when $P(Y) \neq 0$.

Bayes' Rule The probability of one event happening, given that we have knowledge of the occurrence of another event, involving the same variables, is given by the Bayes' rule and is defined as:

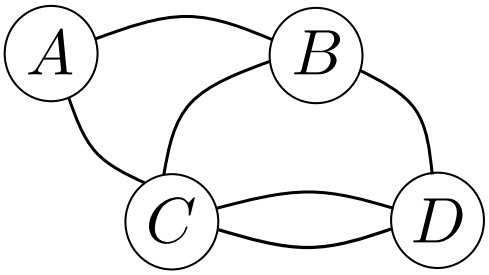
$$P(Y \mid X) = \frac{P(X \mid Y)P(Y)}{P(X)} \tag{3.3}$$

Chain Rule The chain rule, also called the *Product rule*, is used to find the joint distribution of random variables based on their conditional probabilities, deriving from Eqn. 3.3:

$$P(X, Y) = P(X \mid Y)P(X) = P(Y \mid X)P(X) \tag{3.4}$$



(a) A **Directed Graphical Model** showing the interaction between four random variables A, B, C, D. The direction of interaction is statistically important.



(b) An **Undirected Graphical Model** showing the interaction between four random variables A, B, C, D. The direction of interaction is unimportant.

Fig. 3.1.: A graphical model showing the interaction between four random variables classified on the basis of the importance of the directionality.

Joint Probability Decomposition An application of Eqn. 3.4 is the ability to "decompose" a joint probability to simplify its evaluation. Consider a joint probability of three random variables, X, Y and Z, i.e., $P(X, Y, Z)$. This joint probability can be expressed in 13 different, equivalent ways based on the chain rule (Vasquez et al., 2009). The number of factorizations is directly proportional to the number of random variables involved in the Joint Probability. The choice of an appropriate decomposition is dependent on the our knowledge of the probabilistic model. In general, the decomposition chosen is one in which the constituent factors are easily parametrizable or those that reduce the dimensionality of the problem to be solved.

3.2 Graphical Models

In Section 3.1, we gave the example of a three variable system with many different ways of decomposition, with some decompositions being better choices than others. Consider a probabilistic product with another random variable K. Assuming that some of the factorizations in this decomposition involve dependencies on one or more variables jointly or individually for the event to occur, we can safely denote them graphically. Such probabilistic systems are called *Graphical Models*, placed at a conjunction of probability theory and graph theory making them easy to visualise. Complex probabilistic systems and their interactions can be captured and represented by such models as modules and sub-modules. Nodes in the graph represent the different variables of the system and the edges represent the interactions between them.

Depending on the nature of the edges of the graph, these models can be classified into two - directed graphical models and undirected graphical models as shown in Fig. 3.1. If the

directionality of interaction is important in the representation of the model, such a model is called a directed graphical model or an *acyclic directed model*. The edges attain statistical significance. The joint probability of such a system depends on the conditional probability of the parents of each of its nodes $X_v, v \in V$ while the descendants of v have no effect on it. Mathematically, this can be generalised as (Madigan et al., 1995):

$$P(V) = \prod_{v \in V} P(v \mid \mathrm{pa(v)}) \tag{3.5}$$

where $\mathrm{pa(v)}$ denotes the parents of v in the graph.

An undirected graph, as the name suggests, requires no directionality in defining its edges. These models can be decoupled based on the intra-nodal interactions of the model. Consider the undirected graph of Fig. 3.1b. The joint probability of its variables is the product of its conditional probabilities and divided by the probabilities of its independent probabilities. Choosing a directed graphical model to represent a complex interaction becomes evident when the system modelled has a sequential conditional dependence, like representing a time series, for example.

3.3 Inference and Learning

Inference is the main application of probabilistic models. Inference is the process of finding the values of unknown variables based on known ones, generally using Bayes' rule (Eqn. 3.3).

Consider also a sequence of observations captured by a sensor over time. This temporal sequence of data can be represented as

$$O_{t_1:t_2} = \{o_{t_1}, o_{t_1+1}, o_{t_1+2}, \ldots, o_{t_2}\}$$

where the data was captured between times t_1 and t_2, $t_1 \leq t_2$. A probabilistic model simply maps the observed values (*the evidence*) to the most probable cause of this evidence given the parameters of the model. Formally, given the evidence O, the hypothesis X consisting of state variables and parameters θ, inference is the question $P(X \mid O)$ which can be expressed in simpler terms using the Bayes' rule and the product rule. This expression turns out to be:

$$P(X \mid O) = \frac{\sum_{X_k} P(O, X_k, X)}{\sum_{X_k, O} P(O, X_k, X)} \tag{3.6}$$

where X_k is the vector of states uninvolved in the generation of the observation and can be marginalised out. The solution to Eqn. 3.6 is exponential to the number of variables. Thus, for complex probabilistic models, the solution becomes intractable rapidly. One method of

handling this scenario is to apply approximate inference instead. This can be performed in many ways.

One method is to utilise graphical models to model the interplay between the various variables. By exploiting this graphical model for conditional independence and, consequently, theJPD, we can arrive at a simpler expression thus rendering the solution tractable. The effect of such a conversion is to sacrifice accuracy for reduced complexity.

A second method for reducing complexity of a probabilistic model and perform approximate inference is to substitute some of the factorizations of a JPD with elementary distributions such as the *Gaussian Distribution*. The Gaussian, a multivariate normal, can be parametrized by its mean and covariance. In its parametric form, it is expressed as:

$$P(\mathbf{X} = x_i) = \mathrm{G}(\mathrm{x_i}; \mu, \Sigma) \tag{3.7}$$

$$= \det(2\pi\Sigma)^{-\frac{1}{2}} \exp\{-\frac{1}{2}(\mathbf{X} - \mu)^\top \Sigma^{-1}(\mathbf{X} - \mu)\} \tag{3.8}$$

where μ is the mean vector and Σ the covariance matrix and $\mathbf{X}$ the entire domain of the random variable. Generally speaking, graphical models and elementary distribution substitutions are used in tandem to find tractable solutions to probabilistic problems.

In performing such approximate inference, we are led to another problem - that of choosing the right values for the myriad parameters of the constituent distributions for our model. These can either be handcrafted or learnt from available data.

A way of learning these parameters, and the simplest, is to *count* the number of instances the event takes place. This is the *Maximum Likelihood (ML)* estimation of the parameter.

For discrete variables, the application of the ML method to learn the parameters is straightforward. Yet, consider a set of data where there is no instance of an event occurring. A zero occurrence event would lead to undefined values for associated parameters, thereby affecting the accuracy of inference.

Sometimes, data regarding some of the variables in our model is hard to directly observe i.e., they are *hidden*. A common method to estimate this hidden variable is to assign a rational value to it and observe if the effects of this data matches actual observations and to re-estimate it from that point. This is the basis of the *Expectation Maximisation (EM) Algorithm* (Dempster et al., 1977).

3.4 The Bayes Filter

An easy method of performing inference based on hidden variables and observations is the Bayes filter. In essence, a Bayes filter is called so because it uses repeated observations from many sequences to "filter" out outliers and modify the uncertainty states. A direct application of the Bayes filter, in our context, could be to reduce the uncertainty of a pedestrian's

spatial position over the course of an observation sequence. The Bayes filter takes into account model incompleteness and noisy observations. This makes it a prime candidate to answer the prediction question - knowing with a finite valued certainty the spatio-temporal position of an object, what will be the state of this object H timesteps in the future, given that it has been observed in a sequence $O_{1:t}$ until the current time t.

The Bayes filter is composed of two variables -

- S_t, the state of the environment at time t and
- O_t, the observation captured by sensors at time t.

The Bayes Filter is, at its base, a JPD of the two variables defined above. The Bayes filter also defines the following constraints:

- Every observation at a specific time is dependent only on the state that is generating it. This implies that the observation O_t is independent of all other previous observations and all other generative states.

$$P(O_t \mid O_{1:t1}S_{1:t}) = P(O_t \mid S_t) \tag{3.9}$$

- State evolution follows the *Markov Property*.

$$P(S_t \mid S_{1:t-1}) = \begin{cases} P(S_1) & \text{if } t = 1 \\ P(S_t \mid S_{t-1}) & \text{otherwise} \end{cases} \tag{3.10}$$

Eqn 3.9 describes the *Observation Probability* or the Sensor Model. This is the probability that accounts for sensor noise and models the probability of obtaining a specific measurement at a specific state.

Eqn 3.10 is called the *Transition Probability*. This probability assigns a value on how probable it is to achieve another state from the current state.

The Bayes filter has many specializations based on the characteristics of the environment it is to be applied in. A Kalman filter is a specialization that describes an environment wherein the evolution of states and observations are continuous. A graphical variant of the Bayes filter is the Hidden Markov Model (HMM) which will be presented next.

3.5 Hidden Markov Models

A Hidden Markov Model (HMM) is a graphical representation of the Bayes filter with the states being discrete. No such restrictions apply for the observations, however, which can be either continuous or discrete. Here, we provide a short recap on the theory of HMMs

to familiarise the reader on topics of parameter learning and predicting patterns based on observed data.

3.5.1 Probabilistic Model

An HMM is a Markov Chain that emits observations consisting of a finite number of complete, hidden states. Given that the states of the HMM are theoretically *complete*, every state in the chain is independent of the emitted observations and of all other states except the preceding one. The probabilistic model of the HMM is the same as that of a Bayes filter Section 3.4 which allows us to define the JPD as:

$$P(S_{0:T}, O_{1:T}) = P(S_0) \prod_{t=1}^{T} P(S_t \mid S_{t-1}) P(O_t \mid S_t) \tag{3.11}$$

where,

- The states S_t, S_{t-1} are the current and previous states; $S_t, S_{t-1} \in \{1, 2, \ldots, N\}$, a discrete set of possible states of the system.
- O_t is the current observation. This observation can either be discrete or continuous. Generally, $O_t \in \mathbb{R}^k$, where k is the dimension of the state space.

Given all the details on the observations $O_{1:t}$, the distribution of the state S_t can be found by Bayesian reasoning. By recursively defining $P(S_{t-1}) = P(S_{t-1} \mid O_{1:t-1})$, i.e., the posteriors of the previous step becoming the priors of the next time step, eqn. (3.11) can be simplified and re-written as :

$$P(S_{t-1} S_t O_t) = P(S_{t-1}) P(S_t \mid S_{t-1}) P(O_t \mid S_t) \tag{3.12}$$

This decomposition (Eqn. 3.12) contains three distinct models that can be variably defined:

- *Prior State Distribution* $P(S_0 = k) = \pi_k$. This is the state prior vector, the starting point from which the model evolves. $\mathbf{S_0}$ is a vector of size H, where H is the total number of states in the model. $\pi_k \in \{\pi_1, \ldots, \pi_H\}$
- *Transition Model* $P([S_t = i] \mid [S_{t-1} = j]) = a_{ij}$. This model describes the probability of entering a certain state by knowing the previous state. For a stationary HMM, i.e. a time invariant HMM, the transition distribution is defined by an $H \times H$ matrix and represented by $\mathbf{A}$.
- *Emission Model* For a stationary HMM, the emission distribution, also called the observation model, $P(O_t \mid S_t)$ is defined by a matrix $\mathbf{B}$ of size $V \times H$, where V is

the number of observations in the sequence. The parametric form of this distribution is dependent on the nature of the observations. Given a continuous observation, the observation model generally takes the form of a multivariate normal (Gaussian Distribution) defined by its mean μ and its covariance σ.

The final model of the HMM is represented as

$$\lambda = \{\pi, \mathbf{A}, \mathbf{B}\} \tag{3.13}$$

3.5.2 Inference

Inference is one of the main applications of an HMM, having been used for speech recognition (Rabiner and Biing-Hwang Juang, 1993) or action recognition (Lv and Nevatia, 2006). The main inference problems that the HMMs are used to solve are:

1. *State Estimation* Estimation is the process of inferring the belief over the current hidden state of the HMM given the observation sequence $O_{1:t}$. More formally, the application of state estimation can be described as $P(S_t \mid O_{1:t})$. This probability can be calculated recursively as:

$$P(S_t \mid O_{1:t}) = \frac{1}{Z} \sum_{S_{t-1}} [P(S_t \mid S_{t-1}) P(S_{t-1} \mid O_{1:t})] \tag{3.14}$$

   where Z is a normalising factor.

2. *Smoothing* Smoothing of a signal is one the more common applications of HMMs in the field of signal processing and speech recognition. Smoothing is the process in which the current sequence undergoes a "backtracking" and a "forwarding" from that point to "smooth" the hidden state sequence. Thus, smoothing can lead to better *State Estimation*. Formally, smoothing is defined as:

$$P([S_t = i] \mid O_{1:T}) = \frac{1}{Z} \boldsymbol{\alpha_t}(i) \boldsymbol{\beta_t}(i) \tag{3.15}$$

   The terms $\boldsymbol{\alpha_t}$ and $\boldsymbol{\beta_t}$ are the forward and backward probabilities respectively, whose calculation is described in Algorithms (1) and (2) while Z is a normalising factor as in *State Estimation*

3. *Most Probable State Sequence* The Most Probable State Sequence is a consequence of the Viterbi algorithm (Viterbi, 1967) as an answer to the question "Given this observation, what is the sequence of hidden states that led to this observation?". The Viterbi algorithm is similar to the Forward algorithm (Algorithm (1)).

4. *State Prediction* This is the more important of the applications of HMM inference within the confines of our context. Given an HMM, state prediction is the answer to the question $P(S_{t+H} \mid O_{1:t})$, $H > 0$ with a sequence of observations $O_{1:t}$, which can be explained as the propagation of the belief of the State a la *State Estimation*, for H timesteps in the future. This distribution over states can be computed as:

$$P(S_{t+H} \mid O_{1:t}) = \sum_{S_{t+H-1}} [P(S_{t+H} \mid S_{t+H-1}) P(S_{t+H-1} \mid O_{1:t})] \tag{3.16}$$

This recursive computation over all states and over all timesteps from t until H is of a complexity $\mathcal{O}(HN^2)$, given N possible transitions on each state. It is evident that prediction becomes complex at long time horizons.

3.5.3 Learning

Learning in an HMM is equivalent to searching for its most appropriate parameters to approximately explain the recorded observations. More formally, knowing the observation set $\mathcal{V} = \{\mathbf{v}^1, \ldots, \mathbf{v}^N\}$ where $\mathbf{v}^n = O^n_{1:T_k}$, observations of length k, independently generated, the problem of learning an HMM is to find the model λ. This investigation involves the searching for an appropriate transition matrix $\mathbf{A}$, an emission matrix $\mathbf{B}$ and the vector π.

A straightforward method of learning all parameters in graphical models was introduced in §3.3 called the Maximum Likelihood method. This, as previously mentioned, involves counting events at each state for each of the observation sequences in the observation set. This method of estimating parameters cannot be used, however, since the states in an HMM are hidden and only the emitted observations are exposed. Such a case lends itself to the utilisation of the Expectation Maximization algorithm (Dempster et al., 1977) and its derivatives, especially the *Baum-Welch Algorithm* (Baum et al., 1970). This algorithm depends on the the use of the forward and backward probabilities $\boldsymbol{\alpha_t}$ and $\boldsymbol{\beta_t}$ calculated with algorithm 1 and algorithm 2 respectively. The Baum Welch Algorithm has been presented in algorithm 3. For the K observation sequences, the state prior counts, the mean and covariance for the emission probabilities and the transitions between states is estimated based on the calculated forward and backward probabilities as observed. The algorithm presented here is a batch learning algorithm, requiring all K trajectories to be processed together.

For every observation sequence, the parameters in the model λ are recalculated. The Baum Welch Method converges at a local maximum. An issue with the Baum Welch algorithm is that it does not guarantee convergence at the global maximum, instead reaching one of the local maxima depending on the parameters used to initialise the model.

Algorithm 1 Forward Algorithm

Input:
Observation sequence $O_{1:T}$
Model $\lambda = \{\pi, \mathbf{A}, \mathbf{B}\}$
Returns:
Forward Probabilities $\boldsymbol{\alpha_t}(i)$

procedure FORWARD($O_{1:T}$)
  for all i **do**
    $\boldsymbol{\alpha_1}(i) \leftarrow P([S_1 = i])P(O_1 \mid [S_1 = i])$
  end for
  for $t \leftarrow 2 : T$ **do**
    for $j \leftarrow 1 : N$ **do**
      $\boldsymbol{\alpha_t}(j) \leftarrow \left[\sum_{i=1}^{N} \alpha_t(i)P([S_t = j] \mid [S_{t-1} = i])\right] P(O_t \mid [S_t = j])$
    end for
  end for
end procedure

3.6 Discussion

In this chapter, we introduced the reader to the basics of probabilistic models. We also introduced to the reader the Bayes filter and its specialisation, the Hidden Markov Model. We presented the basis of the HMM, its probabilistic model and perspectives on training the HMM and inferring future states based on treated observation sequences. We spent some time in developing the learning algorithm for an HMM which will be used in other parts of this thesis to explain the proposed approach. We also mentioned a certain drawback of the most well known learning process employed in finding the model parameters of the HMM. This is a problem with any EM algorithm used to estimate parameters, the issue of initialising the learning with the right values of the parameters. The Baum Welch method converges quickly, but this convergence can be at one of the myriad local maxima found due to the choices of initial parameters. For this reason, it becomes important to choose logical priors i.e., initial parameters could be chosen via the output of an expert system that understands the task upon which the model is trained.

In general, while HMMs were first used for signal processing and speech recognition (Rabiner and Biing-Hwang Juang, 1993), they have been utilised for diverse applications such as pollution control (Xu and Y. Wang, 2016), to biomedical applications such as prosthesis control (Chan and Englehart, 2005) and gene sequencing (Soruri et al., 2013). Yet, there are drawbacks to this convenient tool. A major drawback is one of topology - an appropriate topology for the task needs to be known beforehand. Methods found in literature focus on a trial-and-error method using the available data to fit a topology (Robles et al., 2012; Vasko et al., 1996). These approaches bring us to another problem - that of data scarcity. The larger the size of the HMM, the more the parameters to be trained and data starvation

Algorithm 2 Backward Algorithm

Input:
Observation sequence $O_{1:T}$
Model $\lambda = \{\pi, \mathbf{A}, \mathbf{B}\}$
Returns:
Backward Probabilities $\boldsymbol{\beta_t}(i)$

procedure BACKWARD($O_{1:T}$)
  for all i **do**
    $\boldsymbol{\beta_T}(i) \leftarrow 1$
  end for
  for $t \leftarrow T-1:1$ **do**
    for $i \leftarrow 1:N$ **do**
      $\boldsymbol{\beta_t}(i) \leftarrow \sum_{j=1}^{N} P([S_{t+1}=j] \mid [S_t=i]) P(O_{t+1} \mid [S_{t+1}=j]) \boldsymbol{\beta_{t+1}}(j)$
    end for
  end for
end procedure

results in a badly trained model. While a smaller model would reduce the complexity, it cannot efficiently capture all observations. On the other hand, having a large model also leads to problems of complexity of time and memory (Khreich et al., 2010).

In the coming chapters, we will discuss solutions to some of the problems presented in this section, specifically the problem of topologies and issues of learning the right parameters.

Algorithm 3 Baum Welch Algorithm

Input

Observation Sequences $O^{1:K}$

Initial values of HMM parameters π_0, a_0

Output

HMM model $\lambda = \{\pi, \mathbf{A}, \mathbf{B}\}$

procedure BAUMWELCH($O^k_{1:T}$)

converged $\leftarrow$ False

while not converged **do**

$\boldsymbol{\alpha}^k_t \leftarrow$ FORWARD(O^k) ▷ Using algorithm 1

$\boldsymbol{\beta}^k_t \leftarrow$ BACKWARD(O^k) ▷ Using algorithm 2

$p_{O^k} \leftarrow probability(O^k)$

for $i \in \{1, \cdots, N\}$ **do**

$\bar{\boldsymbol{\pi}}_{\boldsymbol{i}} \leftarrow \frac{\sum_{k=1}^{K} \frac{1}{p_{O^k}} \boldsymbol{\alpha}_1{}^k(i) \boldsymbol{\beta}_1{}^k(i)}{K}$ ▷ Expected state prior count

$\bar{\boldsymbol{\mu}}_{\boldsymbol{i}} \leftarrow \frac{\sum_{k=1}^{K} \frac{1}{p_{O^k}} \sum_{t=1}^{T_k} \boldsymbol{\alpha}_t{}^k(i) \boldsymbol{\beta}_t{}^k(i) O^k_t}{\sum_{k=1}^{K} \frac{1}{p_{O^k}} \sum_{t=1}^{T_k} \boldsymbol{\alpha}_t{}^k(i) \boldsymbol{\beta}_t{}^k(i)}$ ▷ Observation probability mean

$\bar{\boldsymbol{\sigma}}^{\mathbf{2}}_{\boldsymbol{i}} \leftarrow \frac{\sum_{k=1}^{K} \frac{1}{p_{O^k}} \sum_{t=1}^{T_k} \boldsymbol{\alpha}_t{}^k(i) \boldsymbol{\beta}_t{}^k(i) (O^k_t - \mu_i)^2}{\sum_{k=1}^{K} \frac{1}{p_{O^k}} \sum_{t=1}^{T_k} \boldsymbol{\alpha}_t{}^k(i) \boldsymbol{\beta}_t{}^k(i)}$ ▷ Observation probability covariance

for $i \in \{1, \cdots, N\}$ **do**

$\bar{\boldsymbol{a}}_{\boldsymbol{i,j}} \leftarrow \frac{\sum_{k=1}^{K} \frac{1}{p_{O^k}} \sum_{t=2}^{T_k} \boldsymbol{\alpha}_{t-1}{}^k(i) P([S_t{=}j] | [S_{t-1}{=}i] \lambda) P(O^k_t | [S_t{=}j] \lambda) \boldsymbol{\beta}_{t-1}{}^k(i)}{\sum_{k=1}^{K} \frac{1}{p_{O^k}} \sum_{t=2}^{T_k} \boldsymbol{\alpha}_{t-1}{}^k(i) \boldsymbol{\beta}_{t-1}{}^k(i)}$

▷ Calculating the transition counts between all nodes

end for

end for

if $\lambda = \{\bar{\pi}, \bar{\mathbf{A}}, \bar{\mathbf{B}}\}$ **then**

converged $\leftarrow$ True

else

$\lambda = \{\bar{\pi}, \bar{\mathbf{A}}, \bar{\mathbf{B}}\}$

end if

end while

return λ

end procedure

4 Growing Hidden Markov Models

> *It's all to do with the training: you can do a lot if you're properly trained.*
>
> **— Queen Elizabeth II**

In Section 3.5, we discussed the Bayes filter and its specialization, the Hidden Markov Model. This chapter introduces an extension to the HMM, the Growing Hidden Markov Model (GHMM) method designed to be applied for predicting pedestrian motion. Our contribution in this thesis is an extension on the GHMM method, addressing the shortcomings of the original method to be able to use it in a built urban environment. Here, we present an overview of the GHMM approach and begin with a discussion on GHMM and their working. Consecutive sections of the chapter are devoted to the topics of learning the topologies and parameters of the GHMMs. In section 4.3, we discuss the use of the GHMMs in pedestrian motion prediction. Finally, we conclude in Section 4.4 with a discussion on the shortcomings of the current implementation of the GHMMs.

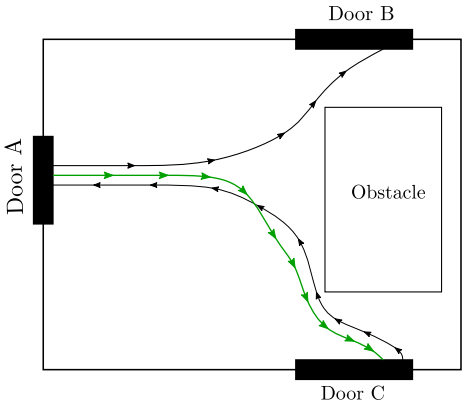
4.1 The Growing Hidden Markov Model

In Section 3.6, we discussed some of the shortcomings of the HMMs in being used for inference applications. The Growing Hidden Markov Model (GHMM) was created as a specialization of HMMs to make the model structure more relevant to the application. This approach was first proposed in Vasquez et al. (2009) to solve the problems of structural limitations of classical HMMs. It was also meant to address their relative rigidity in reacting to new data points as they appear over the course of a time series like that of a pedestrian trajectory.

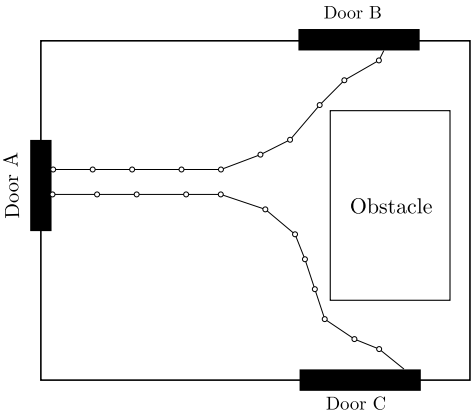
Consider the case of two pedestrian trajectories in a corridor with three doors and an obstacle. One passes the other in the opposite direction with both having destinations as one of the three doors as seen in fig. 4.1. A classical approach to representing these trajectories as a model via HMMs would be to discretize these trajectories at certain intervals as nodes (*States*) connected by links (*transitions*). Such an approach suffers from a few glaring issues.

Model Redundancy Within the corridor, the two trajectories are represented as distinct models with separate transitions and emission probabilities as depicted in Fig. 4.1b. Natural instinct, however, forces us to see that this assumption is not entirely true. Instead, they share a portion of the trajectories with each other which could be unified for representation. Such a unification reduces the overall graphical complexity of the model while increasing the parametric complexity of the model.

Patterning of Trajectories In the classical representation of the trajectories as presented in this case, there is no possibility of transitioning between the two distinct models of the trajectories. Now, consider a new trajectory found in the same set of data of a pedestrian walking from door A to C (shown in green in Fig. 4.1a). This would require another distinct representation while common sense dictates that this new trajectory can be entirely represented piecewise by existing patterns from previous observations. Introducing a separate graphical representation for each new trajectory becomes a cumbersome task and wholly unnecessary and unintuitive. The solution to this problem becomes almost trivial if, at certain junctions, transitions between such close-lying representations are allowed.



(a) Two trajectories starting from Door A to Door B and from Door C and Door A, avoiding the obstacle. A third trajectory starting from Door A and terminating in Door C, in an opposite direction compared to the previously observed trajectory.

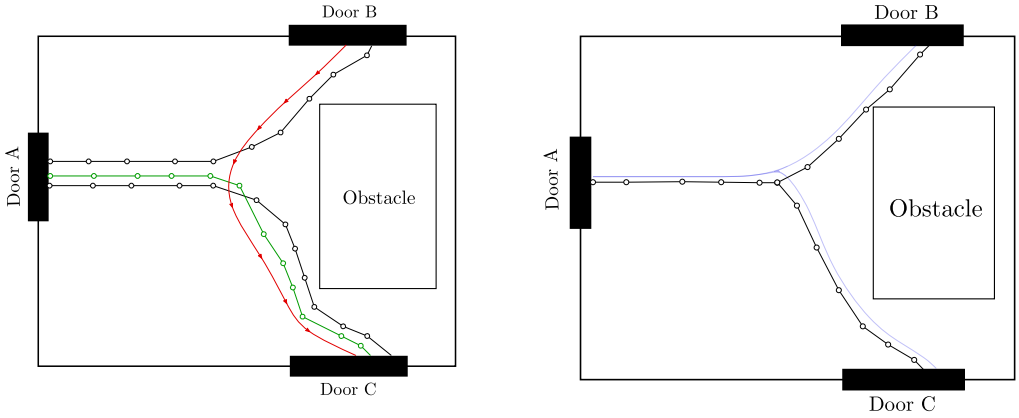


(b) Discretisation of the two observed trajectories into states and transitions between them based on the directionality of motion as modelled for a typical application of an HMM. In this scenario, the green trajectory in the adjacent image cannot be captured by this discretisation and requires a separate one to account for the change in direction.

Fig. 4.1.: Two trajectories close to each other representing typical trajectories within an enclosed space and a possible discretisation of said trajectories.

Anticipation of Trajectories A truly generative model must be able to anticipate a trajectory before such an event occurs. Consider now the case of the same corridor representing

the tracks[1] from door A to B and C. Given that there are sufficient trajectory prototypes contained in the dataset, there is a statistical non-zero probability in imagining a pedestrian path between the doors B and C. Such an exercise in imagining an unobserved specimen of trajectories is a simple exercise to the human brain but is non-trivial to an artificial agent which needs to generate a distinct track for this obvious pattern, a problem mentioned previously. Instead, what is needed is to *anticipate* this track by extrapolating it from existing observations (the red line depicted in fig. 4.2 b.)



(a) With the three trajectories depicted in Fig. 4.1a discretised based on their directionality, the representation still fails to account for a trajectory between doors B and C since a trajectory in this direction between the starting and terminating points have not been previously observed.

(b) A unified graphical representation of the observed trajectories with bidirectional transitions between the nodes of the representation, being able to represent the red trajectory intuitively taken from the adjacent image.

Fig. 4.2.: Given the environment, it should be possible to infer that other, previously unseen trajectories can also be represented with existing models.

Thus the GHMM was developed by Vasquez et al. (2009) to assuage these problems. In essence, the reasoning behind this approach is that any graphical representation of a generative motion model must correspond to the spatial structure of the environment where the events are being observed with the scenario presented in Fig. 4.2b. A meaningful graphical representation must also take into account the vagaries of human motion and the semantic modelling of such trajectories with answers to questions such as "what is he moving towards?" and a rough approximation of the myriad of human intentions possible within an environment.

The GHMM method operates on certain assumptions - a) continuous space is discretisable into a meaningful, finite number of states, b) The state observations are continuous

[1] We use the following convention in distinguishing tracks and trajectories - a trajectory is a directed pedestrian action wherein the path taken by the pedestrian terminates in a destination. On the other hand, we use the term "track" to refer to those parts of graphical representation of pedestrian trajectories which are either non-directed or, more often, bi-directional.

and c) states that are in proximity to one another spatially produce observations that are also close to one another (Vasquez et al., 2009; Thrun et al., 2005).

The GHMM can be thought of as operating in three parallel, multi-dimensional spaces stacked on top of each other, namely the Observation Space, Topological Space and the Model State Space. Interaction between these spaces is unidirectional, as seen in fig. 4.3.

Observation Space The Observation Space $\mathbf{\Omega}$ of the GHMM is the space in which observations on pedestrian movement is captured. The Observation Space is constituted by two overlapping sub-spaces – the positional subspace and the velocity subspace whose components in the observation are vectors continuous in $\mathbb{R}^2$. Observations also point to the intentionality of the pedestrian under observation and an assumption is made that all human motion is goal directed (Hoogendoorn and Bovy, 2004). This allows for the augmentation of the observation space with a *goal* component containing a positional vector in $\mathbb{R}^2$. Thus, every observation[2] $\omega \in \mathbf{\Omega}$ is constituted as:

$$\omega \doteq (\mathbf{x}, \mathbf{v}, \boldsymbol{\phi}) \tag{4.1}$$

where $\mathbf{x}_t, \mathbf{v}_t$ correspond to the instantaneous position and velocity vectors of the pedestrian's trajectory and $\boldsymbol{\phi}$ the intended goal vector of the trajectory. The intended goal is identified post the completion of the trajectory where the last associated observation is designated the goal.

Topological Space The Topological Space $\mathcal{E}$ of the GHMM is a two dimensional Euclidean Space encompassing its nodes and edges as a representation of the graphical model. It is defined by the tuple $\mathcal{E} = \langle \mathcal{U}, \mathcal{L} \rangle$ where N is the number of nodes of the graphical representation and E defines the number of its edges. The Topological Space contains the topological map G of the GHMM – the graphical representation of the environment in which inference is performed. Let $f : \mathbf{\Omega} \rightarrow \mathcal{E}$, then, the function f operates on the topological space, modifying it. More formally, the relation between the topological map G and observations ω is given by eqn.(4.2).

$$G = \{(\omega, f(\omega)) : \omega \in \mathbf{\Omega}\} \tag{4.2}$$

Every observation in a pedestrian trajectory affects the topological map in three ways – a) modification in the number of nodes, b) modification of the number of edges and c) modification of the position of existing nodes.

[2] In this work, O_t is used interchangably with ω. They differ only in the notation that O_t represents an observation at time t while ω corresponds to a general observation with no specific regard to the time at which the observation is captured.

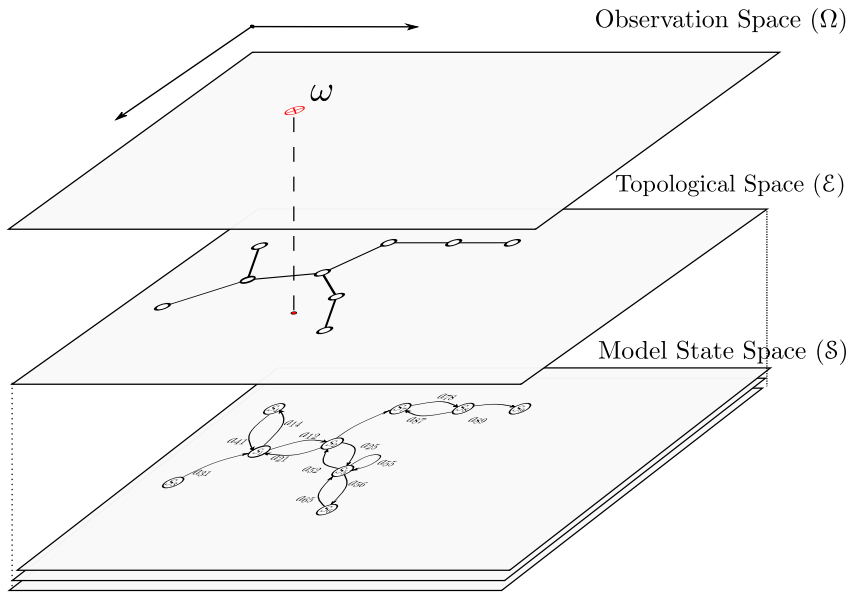


Fig. 4.3.: Spaces of the GHMM

Model State Space The Model State Space $\mathcal{S}$ of the GHMM is defined by the tuple $\mathcal{S} = \langle \boldsymbol{S}, \boldsymbol{\mathcal{A}}, \boldsymbol{S}_0, \boldsymbol{\Phi} \rangle$; S describing the number of states, $\boldsymbol{\mathcal{A}}$ for the number of possible transitions between the states of the the graphical model, $\boldsymbol{S}_0 \in \boldsymbol{S}$, the starting states and $\boldsymbol{\Phi} \in \boldsymbol{S}$, the set of goal states of the graphical model.

In essence, the Model State Space of the GHMM encloses the HMM whose model parameters are defined as $\lambda = \{A, B, \pi\}$. The states of this HMM are derived from the topological map in the topological space $\boldsymbol{\mathcal{E}}$. Given $\{\forall n \in \mathcal{U} | \mathcal{U} \subseteq \boldsymbol{\mathcal{E}}\}$ and $\{\forall \phi \in \Phi | \Phi \subseteq \boldsymbol{\mathcal{E}}\}$, each state $S \in \boldsymbol{S}$ is composed as:

$$S = \langle n, \phi \rangle \tag{4.3}$$

The dimensionality of each S is therefore $\mathbb{R}^4$ since n and ϕ translate to two dimensional real positions on the topological map. Every observation in the observation space is thus incorporated into a state via the topological space $\boldsymbol{\mathcal{E}}$.

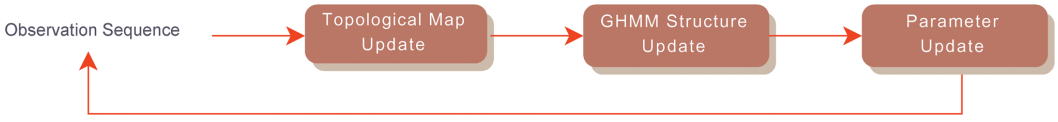


Fig. 4.4.: Original Growing Hidden Markov Model architecture [Image credit: (Vasquez et al., 2009)]

The architecture of the original implementation of the GHMM is presented in Fig. 4.4. This implementation requires a full trajectory for its working. On the completion of a full trajectory, its final positional observation is designated the goal and appended to every ob-

servation in this trajectory to constitute the *extended observations*. These observations are then passed through the pipeline of Fig. 4.4. Each extended observation modifies the topology, thereby *learning* pedestrian patterns to represent their motion. On modification, this topology is translated into a structural change of the graphical model *i.e.,* the underlying HMM. The same trajectory is utilised to learn the different parameters associated with this HMM. In the following sections, we present the method used to learn the topology, structure and parameters of the GHMM.

4.2 Learning a Topology

Any meaningful discretization of the Topological Space $\mathcal{E}$ must imply a logical choice in placing the nodes and the corresponding edges of the resulting topological map G. Following Eqn. 4.2, this becomes a task of defining the function $f(\omega)$. To do so, some assumptions are made:

- *Environmental Conformity* - The topological map G must closely represent the contours of the environment where the observations are recorded.
- *Regional Plasticity* - Nodes discretize the available space and each node defines the center of a plastic region[3] of the topological space.
- *Reachable Topology* - Edges of the graph G, emanating and terminating at nodes can be translated as transitions between the regions whose centers are defined by the nodes. Thus, these edge/transitions should be spatially possible. No edge can exist between two nodes that are not neighbours of each other.

Not only must the function f be able to place nodes and edges on the map G, it must also be able to modify G in order to accommodate new information based on an observation sequence. Such a task requires f to modify positions of existing nodes or even removing them in case of redundancy. Not only does this act on nodes, it should also affect the corresponding edges connecting the nodes.

These criteria can be met by using Topology Representing Networks (TRNs) (Martinetz and Schulten, 1991) which discretize space incrementally and find relations between the regions to connect them. These TRNs are based on the idea of discretizing the space using Voronoi Regions and connecting the centers of those regions with a common border with links using the method of Delaunay Triangulation rendering the centers as nodes and the links as edges of a topological map. The problem now becomes one of setting criteria for the creation of these Voronoi regions. An advantage of using TRNs is that they are easily susceptible to incremental methods of updating the regions.

[3] We define the plasticity of a region as the adaptability of this region, as defined by the center of the corresponding node, to shrink, expand or morph on new observations in the environment.

4.2.1 Instantaneous Topological Map

The Instantaneous Topological Map (ITM) (Jockusch and Ritter, 1999) is a specialisation of the previously mentioned TRNs. The ITM of the GHMM is the function f that incrementally updates the sets of nodes $\mathcal{U}$ and edges $\mathcal{L}$ based on time dependent observations O_t as defined in Eqn. 4.1.

The ITM takes as input the observation and outputs the sets $\mathcal{U}$ and $\mathcal{L}$. It requires the internal parameters τ (the insertion threshold), Σ (the covariance matrix) and $\epsilon \in [0, 1)$ (the smoothing factor).

Every node in the topological map G is associated with a gaussian weight centered on it. The map itself is undirected. Given two nodes i and j, the edges $(i, j) \equiv (j, i)$. As mentioned earlier, a node can only be connected to other nodes if their regions share a common boundary. This leads to the definition of the Neighbourhood of a node i (Vasquez, 2007) -

$$\mathcal{N}(i) = \{j \in \mathcal{U} \mid (i, j) \in \mathcal{L}\} \tag{4.4}$$

The update of the topological map on receiving an observation O_t happens in the following steps -

1. *Node Matching* This is the first step in the ITM process. The two closest nodes based on the the Mahalanobis distance from the observation O_t are found. Since the search complexity is directly dependent on the number of nodes present in the set $\mathcal{U}$, this step can become unwieldy very quickly for large populations of nodes.

2. *Gaussian Adaptation* This step sets the learning rate of the ITM based on the smoothing factor. Every node in the set $\mathcal{U}$ is associated with a bivariate Gaussian. This step determines the rate at which the observation influences the positions (μ) of all the Gaussians.

3. *Node Adaptation* If the observation cannot be associated with any existing node, a new node centered on the observation is created. Since addition of nodes discretise the environment, another node discretising the same space is considered redundant if it is too close to the newly created one, leading to the removal of the older node. A node is also deleted if it is an orphan i.e, there are no edges emanating or terminating in it.

4. *Edge Adaptation* The edge is adapted in two cases - a) if a completely new node is created and b) existing links become redundant in explaining the same motion. If a new node is created in the previous step, then this node is connected to its two closest nodes, maintaining the Delaunay criteria. If the node lies within a multiple of the

insertion threshold, there is a new edge added to this node connecting it to the closest neighbours.

The algorithm itself is as given in algorithm 4 (Vasquez, 2007). The ITM algorithm suffers from a few drawbacks. One of them has already been mentioned - the larger the population size of the nodes, the more complex the computation for finding the closest nodes to an observation. A second drawback is that this implementation of the ITM does not guarantee global convergence. This is due to the fact that the modification of nodes and edges occur locally i.e., on any observation, only the two closest nodes are affected, if any.

4.2.2 Structure Learning

Learning the structure of the underlying HMM in the GHMM is a direct application of the ITM in the state space $\mathcal{S}$. The generation and update of probabilistic graphical model representation of the GHMM is a direct consequence of the observation O_t which updates the ITM. At the end of every observation and the running of the ITM, there is a modification in the sets $\mathcal{U}, \mathcal{L}$. The observation O_t is composed as in Eqn. 4.1. Thus, for every observation, the position and the goal ϕ of the pedestrian being observed is known.

To learn the structure, the following steps are followed:

1. For every node in the topological map, create a state such that it is composed of the node and the goal of the trajectory.
2. If there are edges connecting the nodes in the topology, there are corresponding transitions between these states
3. Each state is associated with certain parameters - the prior value π_0 and a transition probability a_{ij}.
4. On the removal of a node, the corresponding state is removed.
5. On the removal of an edge, the corresponding transitions are set to zero
6. On the addition of a new node, this mode will be assigned a default prior value π_0 and a transition probability a_{ij}.
7. On the addition of every new node, a self transition is added to the corresponding states.

4.3 Learning and Inference with GHMM

The HMM created in the State Space $\mathcal{S}$ as a consequence of the observation sequence $O_{0:T}$ has been described in the previous section. The structure learning of the HMM goes hand

Algorithm 4 Instantaneous Topological Map Update

Input:
Observation O_t
Covariance matrix Σ
Insertion Threshold τ
Smoothing factor ϵ

Output
Modified topological map $\mathcal{E} = \langle \mathcal{U}, \mathcal{L} \rangle$

procedure UPDATETOPOLOGICALMAP($O_t, \Sigma, \tau, \epsilon$)
  $b \leftarrow \arg\min_{i \in \mathcal{U}} d^2_\Sigma(w_i, O_t)$ ▷ Get the closest node to the observation
  $s \leftarrow \arg\min_{i \in \mathcal{U} \setminus b} d^2_\Sigma(w_i, O_t)$ ▷ Get second closest node to observation
  $w_b \leftarrow w_b + \epsilon(O_t - w_b)$
  if $s \notin \mathcal{N}(b)$ and $d^2_\Sigma(W_b, w_s) < 4\tau$ **then**
    $\mathcal{L} \leftarrow \mathcal{L} \cup \{(b, s)\}$ ▷ Edge creation
  end if
  for all $i \in \mathcal{N}(b)$ **do**
    $\bar{w}_{b,i} \leftarrow \frac{(w_i + w_b)}{2}$
    if $d^2_\Sigma(\bar{w}_{b,i}, w_s) < d^2_\Sigma(\bar{w}_b, w_i)$ and $d^2_\Sigma(w_b, w_i) > 1$ **then**
      $\mathcal{L} \leftarrow \mathcal{L} \setminus \{b, s\}$ ▷ Remove this link
      if $\mathcal{N}(i) = \emptyset$ **then**
        $\mathcal{U} \leftarrow \mathcal{U} \setminus i$ ▷ Remove orphaned node
      end if
    end if
  end for
  $\bar{w}_{b,s} \leftarrow \frac{(w_s + w_b)}{2}$
  if $d^2_\Sigma(\bar{w}_{b,s}, w_s) < d^2_\Sigma(\bar{w}_{b,s}, O_t)$ or $d^2_\Sigma(w_s, O_t) > 4\tau$ and $d^2_\Sigma(w_b, O_t) > \tau$ **then**
    $\mathcal{U} \leftarrow \mathcal{U} \cup \{r\}$ ▷ Generate a new node
    $w_r \leftarrow O_t$
    if $d^2_\Sigma(w_r, w_s) > 4\tau$ **then**
      $\mathcal{L} \leftarrow \mathcal{L} \cup \{(b, r)\}$ ▷ Add new link to topology
    end if
    if $d^2_\Sigma(w_b, w_s) < \tau$ **then**
      $\mathcal{U} \leftarrow \mathcal{U} \setminus s$ ▷ Remove redundant node
    end if
  end if
end procedure

in hand with learning the parameters that define the HMM. In Section 3.5, we discussed the Baum Welch method of learning the different HMM parameters. With the insight that pedestrians tend to follow paths of previous trajectories, there is a need for incremental learning of transitions between states i.e., after every trajectory instead of batch learning many trajectories at once. The GHMM method indulges this incremental parameter learning using the Incremental Baum-Welch Method (IBW) (Neal and Hinton, 1998). The HMM model λ remains the same as previously described and will be restated here for convenience in Eqn. 4.5, where $\boldsymbol{\pi}$ is the initial value of the states of the HMM, $\boldsymbol{A}$ is the transition matrix of the HMM and $\boldsymbol{B}$ the emission matrix. The task of learning is to find the appropriate values for these model parameters.

$$\lambda = \{\boldsymbol{\pi}, \boldsymbol{A}, \boldsymbol{B}\} \tag{4.5}$$

4.3.1 Probabilistic Model

Since the GHMM a specialization of an HMM, the probabilistic model remains the same as the latter (as in §3.5.1). Repeating it in Eqn. 4.6:

$$P(S_{t-1}S_tO_t) = P(S_{t-1})P(S_t \mid S_{t-1})P(O_t \mid S_t) \tag{4.6}$$

The JPD can be decomposed as above and each decomposition can be parametrised to solve the inference problem.

- ***State Prior Count*:** The state prior count $\boldsymbol{\pi}$ is the initiator of the IBW. This term is updated after every observation sequence and the resulting value is used as the expectation for the states on the next observation sequence. It can be thought of as the parametrisation of the term $P(S_{t-1})$ in Eqn. 4.6. Whenever a new state is added, this state is initialised with the value π_0.

- ***Transition Probability*:** The transition probability $P(S_t \mid S_{t-1})$ is described by the model parameter $\boldsymbol{A}$. This matrix is initialised by a value a_0 for every viable transition in the model. Analogous to the State prior, this value is updated on every observation sequence. Whenever a new transition is added to the model, it is initialised with a_0. Another thing to node is that on the creation of every state, a self-transition with the aforementioned value is added.

- ***Observation Probability*:** Every node i in the topological map is initialised with a Gaussian $G(\mu_i, \Sigma)$. This Gaussian then is also associated with the state derived from this node, which becomes the observation probability $P(O_t \mid S_t)$. Then, the learning problem becomes finding the parameter mean (μ_i), since the covariance (Σ) is considered to be the same for all nodes.

4.3.2 Parameter Learning

As mentioned previously, the parameters are learnt using the IBW. The parameters π and a_{ij}, the constituents of the model parameter $\boldsymbol{A}$ are calculated via the forward-backward algorithm as the basis of the expectation values of the IBW. The parameter $\boldsymbol{B}$ is learnt when the observation sequences modify the Gaussians associated with the nodes in the topological map. The rate at which this parameter is learnt is modified by changing the **Smoothing factor** (ε).

4.3.3 Inference

We have already spoken about inference in HMMs in subsection 3.5.2. While is some merit to using the underlying HMM of the GHMM to general inference problems like smoothing, we are only interested in its two applications - State Estimation and State Prediction.

Since we deal primarily with pedestrian trajectories, state estimation i.e., the answer to the question "Where is the pedestrian now?" becomes important. The state estimation step in inference can be expressed as below :

$$P(S_t \mid O_{1:t}) = \frac{1}{Z} P(O_t \mid S_t) \sum_{S_{t-1}} [P(S_t \mid S_{t-1}) P(S_{t-1} \mid O_{1:t-1})] \tag{4.7}$$

The second application of inference i.e., state prediction answers the question "Where will the pedestrian be in the future?". The expression describing the evolution of the pedestrian motion is an extension of the state estimation. By propagating the belief held around the states for H timesteps in the future, given that this pedestrian has been observed for t timesteps already can be expressed as in Eqn. 4.8.

$$P(S_{t+H} \mid O_t) = \sum_{S_{t+H-1}} [P(S_{t+H} \mid S_{t+H-1}) P(S_{t+H-1} \mid O_t)] \tag{4.8}$$

Since the state itself is composed of the tuple $\langle n, \phi \rangle$ Eqn. 4.3, it becomes possible to also estimate the goal of the pedestrian over a time horizon by marginalising over the states in favour of the goals. This approach becomes especially useful when the estimating of the pedestrian goals over time becomes important, for example, in the case of a mobile robot wanting to anticipate his actions and reach him.

4.4 Discussion

In this chapter, we have presented a quick primer on HMMs being used to predict human behaviour. We have also given a bird's eye view of the GHMM method and its constituent modules. Specifically, we took an example pedestrian trajectory to explain the building

of the topological map, the changes to the structure and learning the parameters of the GHMM. We outlined the Incremental Baum Welch method for learning the parameters of the underlying HMM. We also explain here the original implementation of GHMMs (Vasquez et al., 2009).

We would, however, like to conclude this chapter with some reflections on the usage of GHMMs for onboard perception and inference for autonomous vehicles. In chapter 2 and in section 4.3, we have discussed existing methods of predicting pedestrian behaviours. As we have seen, methods that discretise their environment - like MDPs and their variants and HMMs - are very popular. A discrete representation of the environment is easier to understand and manipulate. They have also been extensively studied so their use in most prediction applications becomes straightforward. HMMs, especially, have the advantage of being used not only for prediction but also for applications like signal processing (Rabiner, Juang, et al., 1985) and speech-recognition (Rabiner and Biing-Hwang Juang, 1993). GHMMs allow for lifelong, online learning at comparatively low complexity. GHMMs benefit from the depth of knowledge that the domain of HMMs can provide, but also inherit some of the issues that arise with it. In particular, the current implementations of the GHMM suffers from a few drawbacks as can be seen.

These drawbacks can be algorithmic or structural. As an example of an algorithmic drawback, the creation and generation of the topological map becomes expensive based on the number of nodes that already exist in $\mathfrak{U}$. Similar analysis for learning the structure of the HMM can be made and an equivalent insight reached, in terms of complexity.

Environment Agnostic

When talking about prediction of pedestrians, studies (Gibson, 1979; Hine and Russell, 1993; Montel et al., 2013) have shown that the environment is extremely important in understanding pedestrian movement. Other discrete state space models like MDPs found in Ziebart, Maas, et al. (2008) create environmental maps for inference. GHMMs, on the other hand, do not take into account the environment in which the topological map is being built. This oversight can lead to a topological edge connection between two close lying nodes where there should be none. Consider, for example, two separate trajectories in a real environment. Given that these two trajectories don't overlap but move close to each other, two distinct observations could create two separate nodes close enough to be connected by an edge in the topological map. In the real environment, these two observations could be separated by an impermeable wall. This, then, leads to a wrong representation of the environment and thus bad inferences.

Initial Model

The GHMM method utilises the Incremental BaumWelch method (Neal and Hinton, 1998) for learning the parameters of the underlying HMM based on observation sequences. As mentioned in earlier sections, this HMM, described by λ (Eqn. 4.5) of which π is the initial distribution vector. This parameter, also called the state prior, describes the probability of a sequence starting in this state. In training an HMM with the Baum Welch method, we set up the initial model and incrementally update the parameters based on model. Thus, a proper choice of these states are important since the resulting model is heavily dependent on the initial conditions. The Baum Welch method optimises for local maxima and not global maxima. Therefore, the closer the initial conditions are to the actual counts of π, fewer the observation sequences required to reach the optimal values of the parameters. It was found, in N. Liu et al. (2004), that the more complex the HMM structure, the greater the effect of initial parameters. It also showed that initialising the model with parameters derived from real, physical conditions led to better performance. On the contrary, the current GHMM method uses a "State prior counter default value" (π_0), a random value, for initialising any new node that is created. Such an initialisation, with no consideration of the underlying physical condition of the state, results in needing many more observation sequences to attain the global maximum of the environment.

Transition Model

Analogous to the initial model described in the previous subsection, every new node (state) created leads to the addition of transition links between them and existing nodes. These transition links are associated with a transition probability, the "Transition counter default value" (a_0). This value is assigned randomly as with π_0. Considering the example presented in sec. 4.4, assigning a random transition value to two nodes that cannot connect physically in the real world leads to errors in inference. A corresponding case can be made for the issue of the observation probability. In the classical implementation of the GHMM, it is assumed that there is no difference in the covariance of the different Gaussians associated with the nodes of the topological map. This does not ring true in the real world since some nodes can be traversed in only one direction or fewer times than other nodes in the model, thus reducing the probability of observation in certain directions around this node. This distinction cannot be captured within the confines of the current GHMM implementation.

Rich Datasets

The ITM, discussed in section 4.2.1, is the method used to discretise the observed environment. This discretisation is dependent on the observation sequences used for learning the topology and the parameters. A rich and varied dataset is extremely important for the current implementation of the GHMM method. Without diverse trajectories in the dataset,

certain regions of the observed environment could remain undiscretised, resulting in areas where pedestrian motion cannot be predicted. This poses a challenge for the autonomous vehicle relying on this prediction for decision making.

A second reason for requiring a rich dataset has already been mentioned in Section 4.4. More and more full trajectories need to be observed to update the initial model to account for real behaviour in the environment. These reasons also expose a flaw of the current implementation of the GHMM method. It cannot be used in a previously unobserved, novel environment and be used for inferring pedestrian positions. Finally, another need for a rich dataset is the requirement of the *Augmented Observation* as previously mentioned in §4.1. Each observation in the sequence is added with a goal component with a definition that the goal of the trajectory is the final observation of the sequence. Thus, in the absence of a terminating observation in the sequence, the GHMM method fails.

Thus, as we have discussed, there is an opening for improving the GHMM method. Such an improvement must:

- model the environment in such a way as to account for the physical conditions of the discretised environment,
- capture natural human motion around nodes to fit data,
- provide a realistic initialisation and transition model for the learning method to exploit,
- rely on fewer observations without conceding the accuracy of predictions, and,
- be possible to use in a previously unobserved, novel environment to account for AVs motion

In the following chapters, we propose changes to the existing GHMM implementation, specific to the application of pedestrian motion prediction, to solve the drawbacks discussed in this section.

Part II

Modelling And Prediction In Human Populated Urban Areas

5 Modelling the Human Environment

> *We have no idea about the 'real' nature of things … The function of modeling is to arrive at descriptions which are useful.*
>
> **— Richard Bandler and John Grinder**
> Authors

The epigraph of this chapter captures a fascinating idea - that the subtlety and range of human motions cannot be completely understood and that true intentions of actions are frequently not visible. This "hidden" intention, in the context of pedestrian prediction, is the destination of the pedestrian under our consideration. While the estimation of the internal intention of every person is beyond the scope of current technology, falling into the realm of science fiction of needing to read minds, an approximate estimation of these intents can surely be constructed. Such a thought has driven research for years, with literature on pedestrian motion considering his internal intention as a " hidden variable" to be marginalised in the solution (Kautz and Allen, 1986; Bandyopadhyay et al., 2013; Schneider and Dariu M. Gavrila, 2013). On the other hand, another school of thought exists that prefers to explain human behaviour as a function of their spatial constraints (Barker and Wright, 1954; Cannon, 2013). By applying the wisdom present in the epigraph, a useful approximation of intentions can also be built also by externalising the influences on them. Subsequently, in this chapter, we present an approach to model a structured, built environment from the perspective of a pedestrian.

We begin this chapter by discussing some of the different approaches found in literature on the topic of environmental modelling and cost map generation. In Section 5.2, we discuss the aforementioned effect of the built environment on influencing pedestrian behaviour and introduce the concept of Point of Interest (POI). Here, we will further discuss the effects of these POI on pedestrians in urban areas with a firm sociological foundation.

Section 5.3 will deal with the theoretical basis on modelling the various features, identified by common semantic labels, that can be observed in a built urban environment. Next, Section 5.4 details the comparative validation of the proposed approach with a publicly

available dataset and a discussion on the results in conformity with the "Look-Once" principle (Section 1.2). Finally, we present concluding remarks in Section 5.5.

5.1 Advances in Environmental Modelling

A naive method of modelling the observed environment is an occupancy grid map (Moravec and Elfes, 1985). Every cell in such a grid map is defined by it either being occupied or non occupied, creating a static, binary representation of the observed environment. Such a representation is not always sufficient. In fact, it does not take into account the dynamicity of the environment. Many obstacles, that are stationary at point in time, move. This semantic change of occupancy from occupied to unoccupied needs to be captured. One solution is to model this change probabilistically. Such a mapping of these occupied areas was first treated in Thrun et al. (2005). An improved version of these Occupancy Grid Maps was proposed in Weiss et al. (2007), wherein a binary Bayes filter is used to divide the cells on the map into moving and stationary obstacles and estimating the velocities of each of these cells. Other approaches exist, like Rummelhard et al. (2015) which utilise Bayesian approaches to model dense occupancy of the observed area. The simple modelling of occupancy is a reactive measure, insufficient to explain many of the interactions within the environment. Some approaches like Wolf and Burdick (2008) model the interactions in the environment, specifically for highway scenarios while Svenstrup et al. (2010) approaches the problem from a similar direction, modelling the environment as a sum of its interactions. These models are then used to plan agent trajectories in indoor or outdoor scenarios.

Consider a typical urban scene (Fig. 5.1). A pedestrian stands on a side-walk, close to a cross-walk, intending to cross to the other side of the street. A car is moving in a direction perpendicular to his crossing. Another car is moving in the opposite direction, closer to the cross-walk. Consider also, an observer $\mathcal{M}$ witnessing this scene and wanting to predict this pedestrian's behaviour in this situation. Following the approach of Weiss et al. (2007) applied to the problem, $\mathcal{M}$ will not consider the presence of the cross-walk. Comparably, the approach of Rummelhard et al. (2015), fails to include pedestrian intent in that he can stop and wait for the car to pass him by before taking action.

Thus, as an indicative measure, the above mentioned approaches suffer from the misexplanantion of the following problems:

1. **Configuration**: In building an occupancy map, the underlying configuration[1] of the environment is ignored.

[1] Configuration, in this context, is the collection of the different elements of the environment such as grass, pavement, road etc.

2. **Kinematics**: These approaches do not take into account the dynamic and kinematic constraints of the obstacles that it is trying to model. For example, humans - who are dynamic obstacles - can change their speed and direction on a whim.

3. **Context**: The richness of information present in the interactions between the different elements of the scene is lost.

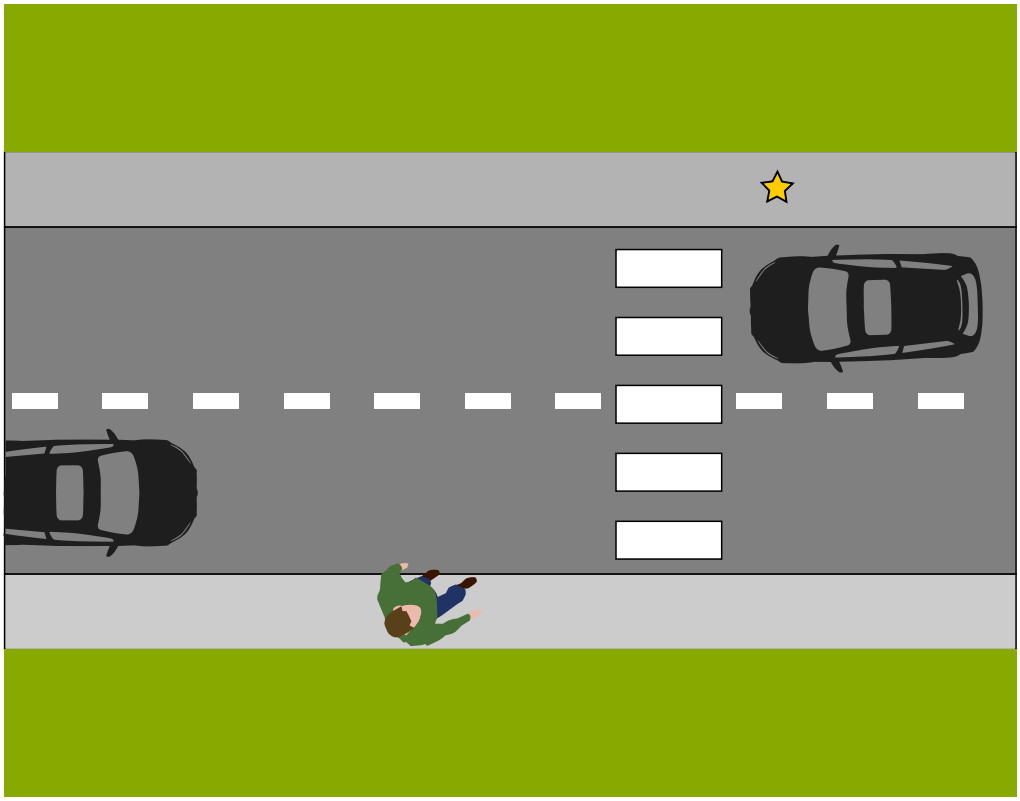


Fig. 5.1.: A typical urban scene with a cross-walk, side-walks and cars sharing space with pedestrians. A pedestrian wishes to cross the road to reach his destination, denoted by the star in gold.

The problem of **Configuration** has been studied with some success. Data driven approaches have been used to identify configurations of the observed space which are more conducive to human interaction. In Ziebart, Ratliff, et al. (2009), a static, closed area is considered for modelling. The area is discretised and pedestrian trajectories through it are observed. A global model of the environment is built such that the cost-to-go to each cell from any other cell of the discretised map, can best explain the set of observed trajectories. This is performed using maximum entropy inverse reinforcement learning (Ziebart, Maas, et al., 2008). As an intuition, this map then denotes the areas which are easily traversable for a pedestrian and which areas are too "costly" to walk in.

An approach to the second problem, i.e. that of **Kinematics**, is seen in Kanda et al. (2009), which utilises the velocities of observed pedestrians in an environment to identify

areas of global behaviour. Sensors pick up pedestrian movements. The environment is discretised into Voronoi regions and the velocities of each observed pedestrians within each of these regions are recorded. An SVM classifier is used to classify behaviour - idle walk, brisk walk, stop, wandering - in each region of the environment, resulting in a global map of pedestrian behaviour there. In a related approach, Z. Wang, Jensfelt, et al. (2016) utilize a probabilistic model to capture pedestrian behaviour in the environment using a non-ergodic HMM. Each cell in the grid is 8 - connected and pedestrians can move in one of the eight directions. This movement corresponds to a transition from one state to another in the corresponding HMM. After every observation, these transitions are updated and a typical "behaviour" is learnt. These behaviours are clustered and classes for each cell is obtained, leading to a grid map with identifiable behaviour probabilities.

A hybrid approach, expounded by Kitani et al. (2012), builds on the work done in Ziebart, Ratliff, et al. (2009) and also manages to take into account the semantics of the observed scene. The environment plays a large role in deciding the trajectory of a pedestrian. For example, going back to the scenario described in Fig. 5.1, the pedestrian is far more likely to cross on the cross-walk to get to his destination rather than cross on the street. He is also likely to avoid the car in his crossing. This interplay between the agent's trajectory and the physical properties of the scene are considered. This approach is unique in comparison with the other approaches discussed here since the learning is transferable. Weights learnt in one scene for features with many pedestrian trajectories are associated with the underlying physical, i.e., semantic, information and not the location. This implies that parameters learnt on one scene can be transferred to a novel scene and begin to infer pedestrian positions even with no observations. Another approach that utilises similar principles can be found in Vasquez (2016), which uses the configuration aware map presented by Ziebart, Ratliff, et al. (2009) with a focus on the kinematics of the pedestrian. A comparison between the different approaches can be seen in Table 5.1. While certainly not an exhaustive list by any means, it can be indicative and a clear picture emerges from Table 5.1. Of the different approaches presented, very few take into account the problem of **Context** while modelling the environment.

	Configuration	**Kinematics**	**Context**
(Ziebart, Ratliff, et al., 2009)	✓	✓	×
(Kitani et al., 2012)	✓	×	✓
(Vasquez, 2016)	✓	✓	✓
(Z. Wang, Jensfelt, et al., 2016)	✓	✓	×
(Kanda et al., 2009)	×	✓	×

Tab. 5.1.: Comparison between the works found in literature on the basis of the problems of Configuration, Kinematics and Context in modelling. Some problems are solved during the planning phase and not during the modelling phase.

It also soon becomes evident that there are some drawbacks inherent with the works reviewed here.

1. There is a need for many trajectories to accurately model the environment. Fewer trajectories lead to an incomplete model of the environment.

2. Strange behaviours and previously unseen behaviours cannot be accurately accounted for.

3. Weights learnt at one scene cannot be generalised without modifications and do not consider dynamicity of obstacles.

4. Cues from Context are unaccounted for.

In general, due to the above mentioned drawbacks, these approaches cannot be utilised by the observer $\mathcal{M}$ on a moving platform like an autonomous vehicle to model a rapidly changing environment.

Human drivers, on the contrary, follow the "Look-Once" principle where they, metaphorically, put themselves in the pedestrian's shoes and predict what they might do in this situation. Thus, any approach that proffers to replace a human observer in the self-driving pipeline must be able, as soon as the area is observed, to : (a) create a mental cost map of the environment; (b) identify areas of high pedestrian distribution; (c) distinguish between possible destinations.

In the following sections, we present an approach that considers the above ideas to formalise a numerically exploitable model using the different features of the observed environment. Our aim is to constitute a function $J(\cdot)$ for the observer $\mathcal{M}$ taking into account the configuration, context and dynamics of the different elements of the observed environment. In the next section, we present the influences on pedestrian behaviour in a built, urban area.

5.2 Modelling What You see: A Sociological Perspective

Humans, as is said, are social creatures. What we see and feel influences our behaviour, especially as pedestrians.

A first approximation of pedestrian motion was proposed by Helbing (1993), called the *Gaskinetic model*. Humans are considered to be gaseous particles and their motion was thought of as being described by Boltzmann-like equations. Pedestrians are considered self-driven particles, moving to unfilled areas within the environment, their motion decided by parameters like velocity and direction plugged into the equation to account for "freedom of choice". An improvement over the *Gaskinetic* equations was the *Social force model*. This widely regarded model was meant to address some of the issues that had arisen with the previous model. Pedestrians are no longer thought of as being gaseous particles. Instead,

they are assumed to have a *social field* around them and all interactions with other elements in the scene happen as a function of this social field (Helbing and Molnar, 1995).

Any model that approximates pedestrian behaviour must also be able to answer the following questions about their trajectories:

Q1 When does a pedestrian decide to move?

Q2 How is the trajectory composed?

Q3 Why does the pedestrian move?

These questions define the idea of Pedestrian Route Choice which plays a major role in understanding their behaviour. Statistically, some behaviours in urban areas can be frequently observed as seen in Hoogendoorn and Bovy (2004). These can become the basis of pedestrian route choice as follows.

1. **Route Directness:** Long term observations of pedestrian behaviour in urban areas show that they are averse to taking long detours. They are also opposed to moving in a direction opposite their desired walking destination. This behaviour has been observed to be true even when their trajectory takes them through crowded areas (Hughes, 2002).

   Contrary to the previous point, there is also sufficient evidence which shows that while pedestrians try to take a direct route, they also prefer taking the fastest route which may not be the shortest one between the origin and the destination (Ganem, 1998). Thus, an inference can be made that all routes are within an area bound by a subjective "fastest" path and a "shortest" path. This polygonal area encompasses all possible trajectories between the current position of the pedestrian and his intended goal.

2. **Proximity:** Pedestrians tend to keep a varying distance between themselves and other elements of the environment, especially with other pedestrians, as explained by the social force model. Pedestrian behaviour is affected by elements like edges of the street, walls etc. (Hoogendoorn and Bovy, 2004; Helbing and Molnar, 1995). Thus pedestrian trajectories suffer a change around walls, side-walks and street edges.

3. **Knowledge:** The knowledge of the environment by the pedestrian plays a huge part in his/her route choice behaviour. Knowing the layout of the street, for example, can influence a pedestrian to choose a shorter crossing, cross where they shouldn't or even force them to avoid a certain part of the road altogether.

Takeaway 1: There are costs manifest in pedestrian routes: both psychological and physical, that are conjoined to the built environment.

5.2.1 Influences on Pedestrian Behaviour : Points of Interest

Continuing with the influences, some areas of the environment are more attractive to a pedestrian than others. To understand this phenomenon and *a priori* identify these areas in a novel environment, we look to the field of cognitive psychology. There are certain *affordances* in the environment that the pedestrian perceives which leads him/her to move in socially significant areas while avoiding others (Gibson, 1979). Intuitively, the affordance associated with an area is the inverse of the cost to traverse that area. This perception of the environment leads to an attractive model of human behaviour - pedestrians move in a direction in free space that allows them to continue moving. This becomes an explicit constraint from the environment. Such directed motion derived from the configuration of environmental elements is defined as ***Natural Vision*** by Gibson (1979). According to him,

> "*When no constraints are put on the visual system, we look around, walk up to something interesting and move around it so as to see it from all sides, and go from one vista to another. That is natural vision ..." (1979, Page 1)*

Thus, in a typical urban scene, we infer on reasoning, that a combination of various factors affect trajectories. Typical pedestrian behaviour, at a high level of reasoning, is a sequence of actions such as "look around", "find most desirable route" and "move in that direction." This *natural* behaviour is explored in Hillier et al. (1993) leading to the theory of *Natural Movement*. In Hillier et al. (1993) it is exhibited that most pedestrian interactions and movements occur along lines of sight. These lines of sight facilitate human movement along them. Hillier et al., contrary to previous definitions, consider these converging lines to be the "configuration" of the environment. The termination points of these lines of sight are designated attractors - positive or negative.

Positive Attractors Some areas in the field of view act as a positive influence on the trajectory of a pedestrian. Take, for example, a monument of national interest like the Eiffel tower. Most trajectories around the area of the Eiffel tend to converge and terminate at the tower itself. Another common example would be the increased attractiveness of a restaurant near an industrial area during work hours. A large number of people would have their destination as the restaurant around this time. Thus, these areas become statistically a positive attractor in the environment.

Negative Attractors Conversely, there also exist negative attractive elements in the scene. These are those elements that a pedestrian would wish to avoid or walk away from during the course of their trajectory. People, for example, generally walk as far away from a trash bin as possible. If there exists another path in the visual field of the pedestrian, then he/she would prefer to take that path instead (Gibson, 1979).

Thus, these statistically significant positions along the line of sight that can change a pedestrian's behaviour towards themselves become attractors, which we refer to as **Point of Interest (POI)** in the remainder of our work. Some attractors have been described in Gibson (1979), like places of public interest, restaurants etc. while others ((Borgers and Timmermans, 1986), (Pacione, 1980), (Bennison and Davies, 1977)) have performed empirical analyses on pedestrian flow around city centres which have shown that pedestrian route choice is influenced by many commonly observed urban elements like the spatial placements of transportation areas (bus and tram terminals, car parking spaces, etc.), shopping centres, large stores and the environmental linkages between these disparate elements.

Takeaway 2: Sociological attraction points in the environment are recognisable and quantifiable, acting as termination points for pedestrian trajectories.

5.2.2 Pedestrian Crossing Behaviour

While we saw in the previous sections on how environmental configurations and POIs affect pedestrians (regarding **Q1** and **Q3**), we should discuss how these trajectories are composed (**Q2**). Within the context of a built urban environment, this implies searching for explanations on pedestrian crossing behaviour.

We commence by defining this behaviour which can be predominantly divided into *Legal Crossings* and *Illegal Crossings*. Consider once more, Fig. 5.1. In this scene, if the pedestrian decides to traverse the street completely on the designated area, i.e., the cross-walk, this trajectory is considered a Legal Crossing. On the other hand, if he decides to cross the street at any other position, that trajectory can be classified as an Illegal Crossing.

Peponis et al. (1990) essayed to find "patterns" in the negotiated paths in constricted spaces and to formalise them based on empirical observations. The rules can be stated as:

1. Pedestrians avoid unnecessary backtracking.
2. With no distractions, pedestrians continue along the same line of sight.
3. Pedestrians diverge from their current line of sight when a new view allows them more space for activity or a different line of sight.

At a macroscopic level, for a person trying to cross the road, physical and psychological barriers exist. These physical barriers may be actual barriers blocking their way, such as a parked car or a gate. They may also be psychological - an oncoming car that is speeding strongly dissuades a person from attempting a crossing, even if the approaching car is distant. This relationship was studied in Hine and Russell (1993), based on observed data and self reporting of the involved pedestrians. It was concluded that given sufficient barriers, both psychological and physical, the pedestrian might not cross and in some cases, not take that

route at all, preferring an "easier" choice. Similar studies conducted in different parts of the world culminated in equivalent results. In Das et al. (2005), an additional finding of the study was the presence of extrinsic and intrinsic "gaps". The extrinsic gap was the true gap between vehicles that could provide an affordance to pass in between. This extrinsic gap was countermanded by a personal "intrinsic" gap perceived by the pedestrian based on his/her own previous experiences. It was also found that larger the vehicle as a barrier, the greater the hesitation in crossing. In similar vein, Montel et al. (2013) conducted an experiment to determine the effects of different urban configurations on illegal pedestrian crossing decisions. It was found that a major factor in the decision process was the presence and function of the built environment i.e., the presence of shops and pedestrian density. Another key factor was the presence of side walks. A third factor in the decision to cross illegally was the width of the street.

Takeaway 3: The geometric dimensions of the environment play a significant role in the composition of a pedestrian's behaviour.

5.3 Modelling of the Environment as Potential Costs

The aim of this section is to establish the function $J(\cdot)$ that was introduced in an earlier section. Specifically, this function will consider the **Takeaways** presented previously, delving into geometry and the configuration of the environment. We see from the discussion in the previous section that there exist multiple influences on pedestrian behaviour in an urban environment. With the discussion, an overall view begins to emerge - one where pedestrian movement is a function of the built environment with each of the features of the scene affecting it differently.

In terms of crossings, for the occurrence of a legal crossing in a built environment, the following assumptions are made:

- **Road**: The road acts as a barrier for crossing, repelling pedestrians towards the side-walks.
- **Edges**: The edges of the road repel pedestrians such that their paths are restricted to the side-walk.
- **Obstacles**: Static and Dynamic obstacles on the road are repulsive in nature, increasing the resistance of the road and pushing back pedestrians towards side-walks.
- **POI**: Points of Interest are a reason for pedestrians to cross from one side of the street to another.
- **Crosswalk**: A cross-walk acts as a conduit between the two sides of the street and offers no resistance to crossing

- **Sidewalk**: Side-walks offer no resistance to pedestrian movement.

As a consequence, an illegal crossing occurs when at least one of these assumptions is violated.

For each of these assumptions, we can find an analogue in the field of Artificial Potential Field (APF)s (Khatib, 1985) that have been used extensively for modelling the environment for navigation of wheeled robots. APFs can provide a directed cost value at a point in the environment like in Svenstrup et al. (2010) and Wolf and Burdick (2008). What follows is a discussion on translating the geometry and semantic information of the identified features of the scene into costs, thereby modelling the environment as a "Potential Cost Map".

Consider the Observer $\mathcal{M}$ observing a typical urban scene as in Fig. 5.2. This scene can be thought of as existing in two subspaces - the real observation space $\Omega \in \mathbb{R}^2$ and a semantic feature space Ψ^U containing semantic information about the observed scene. Then, there exists a function $h(\cdot)$ that maps the semantic feature space to the real space Ω.

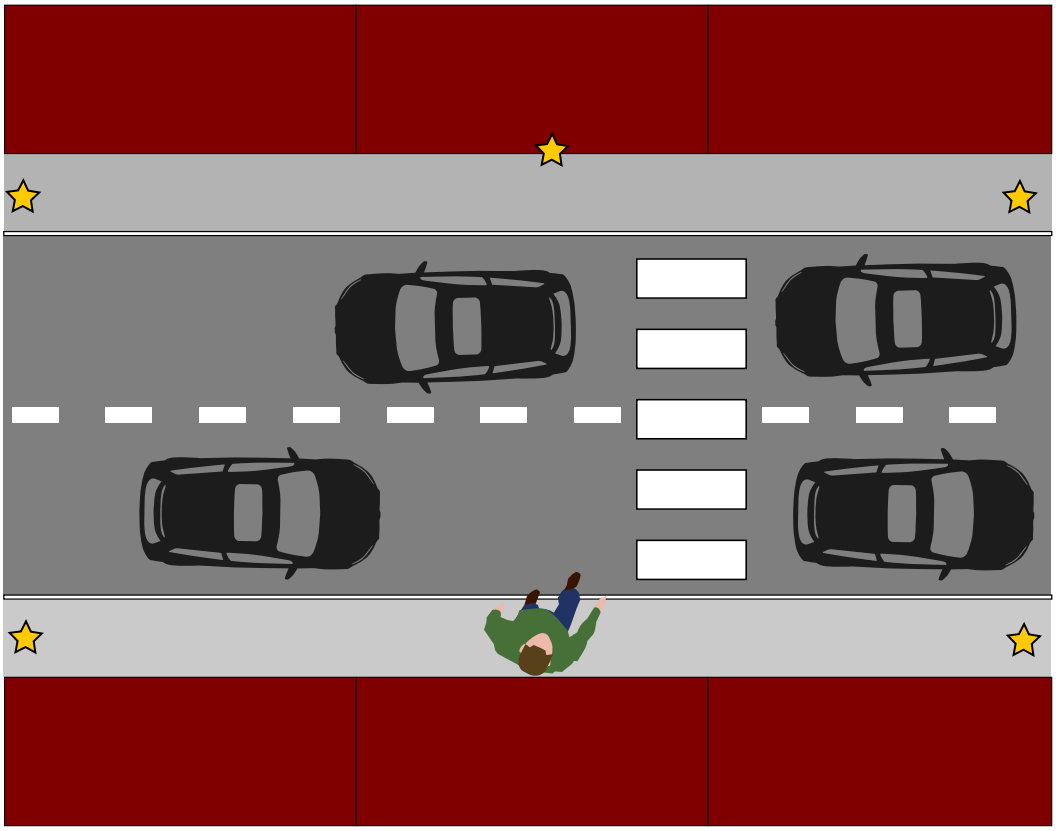


Fig. 5.2.: A typical urban scene containing a pedestrian, points of interest (entrance to a building, edges of the scene), road, sidewalk, crosswalk and dynamic obstacles.

Within this urban scene, we have a feature set $\Psi \subseteq \Psi^U$ such that,

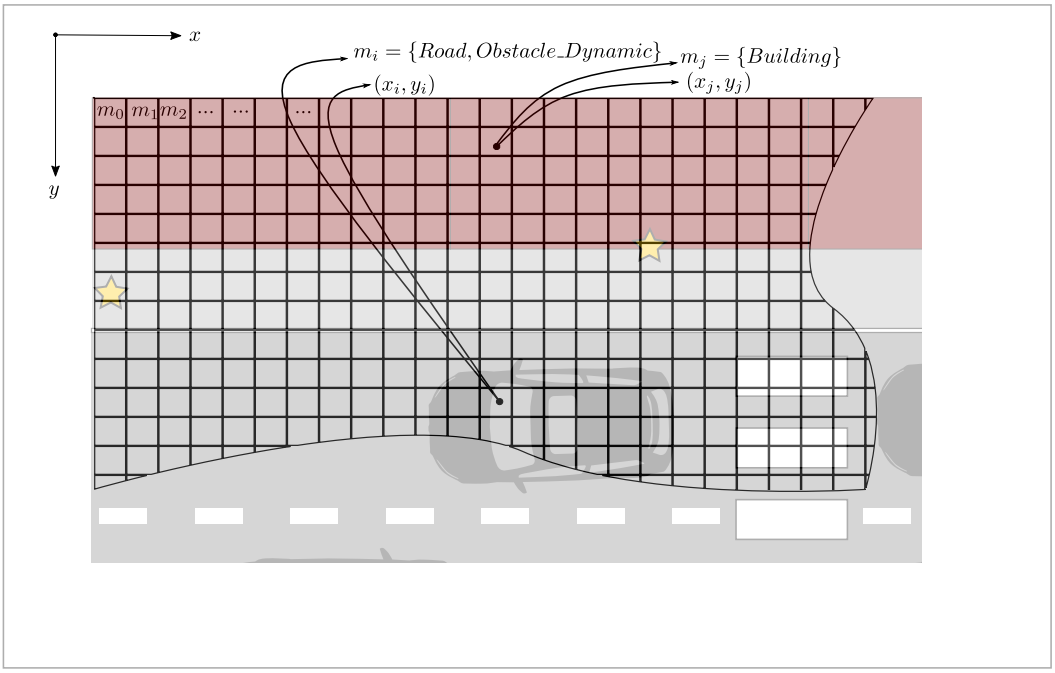


Fig. 5.3.: Semantic grid describing the mapping of Ψ to $\mathbb{R}^2$

$$\Psi = \{Road, Edge, Crosswalk, Sidewalk, O_{Static}, O_{Dynamic}, Building, Point\_of\_Interest\} \quad (5.1)$$

Allowing the observer $\mathcal{M}$ to have the ability to identify these labels, we take advantage of the fact that every point in $\Omega \in \mathbb{R}^2$ is classified into one or more of the labels present in Ψ, as function $h(\cdot)$. Following this, and taking into account the **Takeaways** from the previous section, we can claim that each of these semantic features have a specific influence on a pedestrian, which when taken together, forms his/her trajectory. Thus, $J(\cdot)$ transforms into $J(\Omega; \Psi)$ enabling us to transform the observed features into a cost map.

We begin by discretizing the observed scene into a semantic grid map as shown in fig.5.3. Each cell in the grid, m_i, is defined by its center (x_i, y_i) where $x_i, y_i \in \mathbb{R}^2_{\geq 0}$ and $m_i \subseteq \Psi$ and $i = 0, 1, 2, 3, \ldots \in \mathbb{Z}^+$. With this background, we begin to define specific costs for each of the semantic features as below.

5.3.1 Modelling the Road

As mentioned in section 5.3, the road acts as a barrier for crossing, repelling pedestrians towards the side-walks. Following Wolf and Burdick (2008), the road potential i.e., the cost to cross the road, should be high along the width of the road while tapering off close to its

edges. Thus, it is expected that this feature results in a value $\mathbf{U}_{Road} \in \mathbb{R}$. By defining X as the coordinate values of the observed environment, we have:

$$J(X; [\Psi = Road]) = \mathbf{U}_{Road} \tag{5.2}$$

Returning to Montel et al. (2013), it can be inferred that the geometry of the road present before the pedestrian, especially the width, is a dominant parameter in forcing legal crossings. Intuitively, a small road is more likely to entice illegal crossings rather than a wider road. By this assumption,

Let $X = [x, y]$ defining the x and y the discretisation of the semantic grid.

Then, we can define a Gaussian-like function for cost, c_i, such that,

$\forall i : i \in \mathbb{Z}^+, m_i = Road$,

$$c_{i,Road}(X) = \beta_{Road} \exp\left(-\left[\left(\frac{x_i}{\sigma_k}\right)^2 + \left(\frac{y_i}{\sigma_l}\right)^2\right]\right) \tag{5.3}$$

The total Road Potential Cost, $\mathbf{U}_{Road}$, is:

$$\mathbf{U}_{Road} = \sum_{m_i}^{M} c_{i,Road} \tag{5.4}$$

The Road potential cost for the typical scene depicted in Fig. 5.2 is presented in Fig. 5.4. As can be seen in the figure, the cost of the road is prohibitive enough to cause the pedestrian to move away from the road area. For obtaining this potential cost, the geometry of the road needs to be known *a priori* or estimated to be able to set the values of β_{Road}, σ_k and σ_l. At the end of this step, $\mathbf{U}_{Road}$ is normalised.

5.3.2 Modelling the Edges

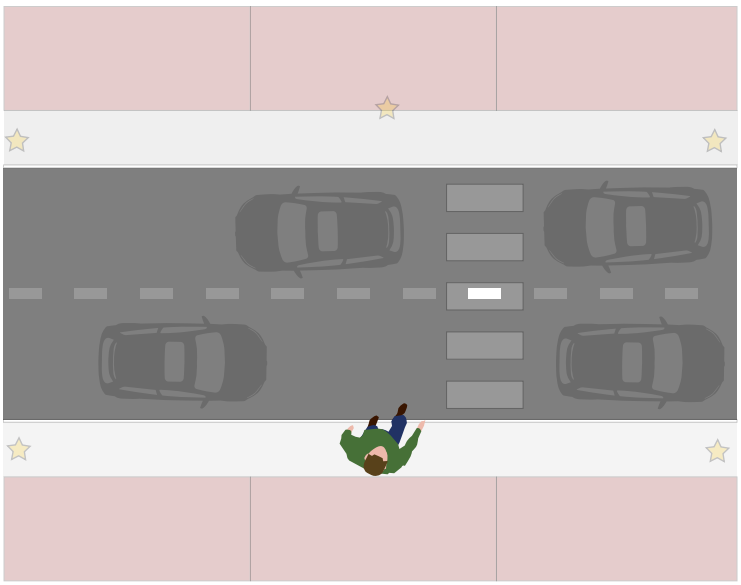
Continuing, we consider the assumption that the edges of the road repel pedestrians towards the side-walk. An illegal crossing occurs when the pedestrian can exert enough force to overcome this potential cost. Similar to Eqn. 5.2, we define the function J for edges as,

$$J(X; [\Psi = Edge]) = \mathbf{U}_{Edge} \tag{5.5}$$

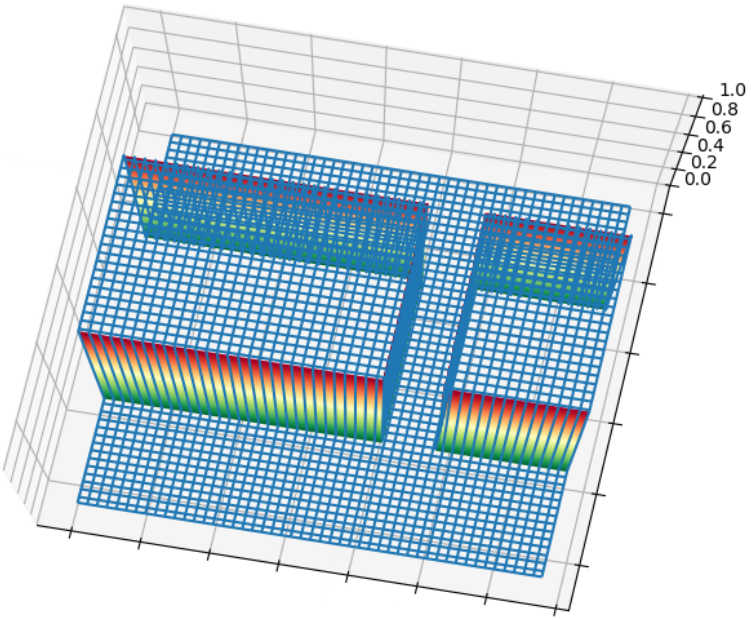
Once again, let X and Y contain the x, y coordinates of all cells of the semantic grid.

$\forall i : i \in \mathbb{Z}^+, m_i = Edge$:

$$c_{i,Edge} = 0.5\eta\kappa(x_i, y_i)^{-1} \tag{5.6a}$$



(a) Highlighted “Road” semantic feature of the typical urban scene.



(b) Potential “Road” Cost for the typical scene depicted. The cost for this feature is dependent on the geometric width of the road.

Fig. 5.4.: Road Potential for the typical urban scene depicted above. The potential is generated for those cells in the *Semantic Grid Map* classified as *Road*.

where η is a scaling factor dependent on L_{Road} and,

$$\kappa = \begin{cases} \rho(x_i, y_i) & if \rho > 0 \\ \kappa_{max} & if \rho = 0 \end{cases} \tag{5.6b}$$

ρ is the distance function given by,

$$\rho = \sqrt{(X - x_i)^2 + (Y - y_i)^2} \tag{5.6c}$$

The total Edge Potential Cost, $\mathbf{U}_{Edge}$, is:

$$\mathbf{U}_{Edge} = \sum_{i \in \mathbb{Z}^+} c_{i,Edge} \tag{5.7}$$

Fig. 5.5 illustrates the effect of the "Edge" feature on a pedestrian in a typical urban scene. Considering that the pedestrian wishes to cross to the other side of the street in our scenario, Fig. 5.5b depicts the costs applied by the environment on him. The edges, as can be seen, have the effect of pushing him away from the edges. An illegal crossing can occur in this scenario when a sufficiently determined pedestrian tries to cross over the edge where the cost is high. As before, the potential cost value for the *Edge* feature is normalised after calculation.

5.3.3 Modelling Obstacles

Obstacles in the scene can be distinguished as static and dynamic obstacles. Pedestrians cannot cross across the obstacle but traverse around them. Pedestrians also respond differently around static and dynamic obstacles, with more care taken to avoid crossings when there are dynamic obstacles in the scene (Montel et al., 2013). Pedestrians approach the obstacle slowly and are pushed away by its edges.

These constraints need to be taken into account to define

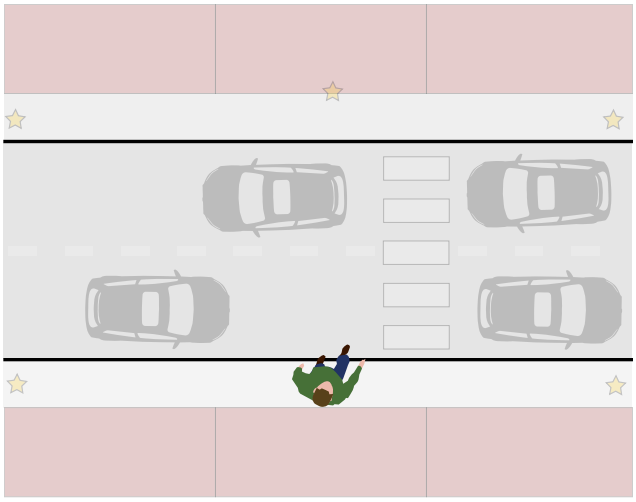
$$J(X; [\Psi = Obs]) = \mathbf{U}_{Obs} \tag{5.8}$$

A Yukawa potential (Volpe and Khosla, 1993) is considered a fit for the expected behaviour. Further, there is difference between the representation of the static and dynamic obstacles for calculating costs. Thus,

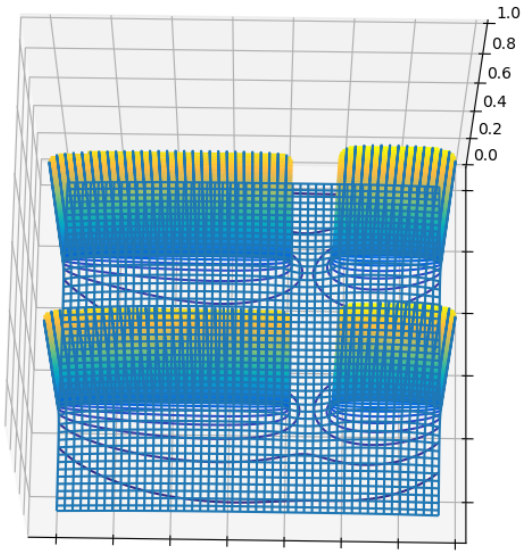
$$J(X; [\Psi = Obs]) = \begin{cases} J([\Psi = O_{Static}]) & \text{if } Obs = O_{Static} \\ J([\Psi = O_{Dynamic}]) & \text{if } Obs = O_{Dynamic} \end{cases} \tag{5.9}$$

and $J(\Psi)$ is applied recursively on every instance of Obs regardless of it being a static or a dynamic obstacle.

Static Obstacle A static obstacle occupies the semantic grid in the shape that it is perceived to be and is thus impermeable. Thus, $J([\Psi = O_{Static}]) = \mathbf{U}_{Stat}$ is defined as,



(a) Highlighted "Edge" semantic feature (in black) of the typical urban scene.



(b) Potential "Edge" Cost for the typical scene depicted above. The lines emanating from the edges show the effect this feature has on a pedestrian, in this case, pushing him away towards the safer sidewalk.

Fig. 5.5.: Edge Potential for the typical urban scene.

$$\mathbf{U}_{Stat} = \begin{cases} u_{max} & \text{if } m_i = O_{Static} \\ 0 & \text{Elsewhere} \end{cases} \tag{5.10}$$

Where u_{max} is the maximum possible value of the potential cost. On normalisation of the total cost (Eqn. 5.18) becomes 1.

Dynamic Obstacle For a dynamic obstacle defined by position and velocity $(x_i, y_i, \dot{x_i}, \dot{y_i})$, we follow the principle presented in Wolf and Burdick (2008), where a triangular appendage on the back of the car generates a cost based on its velocity. On the contrary, we add a triangle to the front of the obstacle in the direction of motion to represent the dynamics and future positions of these obstacles. The vertex of the added triangle is placed based on the velocity of the object itself i.e., the position where the center of the obstacle will be for a certain horizon given that the car is travelling at an observed velocity.

The Obstacle Potential cost $J([\Psi = O_{Dynamic}]) = \mathbf{U}_{Dyn}$ for the n^{th} dynamic obstacle is given by:

$$\mathbf{U}^n_{Dyn} = \Lambda \frac{\exp(-\alpha \mathbf{K})}{\mathbf{K}} \tag{5.11}$$

Where Λ and α decide the behaviour of U^n_{Dyn}. Larger the values, sharper the drop off of the potential near the obstacle.

$\mathbf{K}$ is the distance of the obstacle from every point on the workspace, i.e.,

$$\mathbf{K} = \|C^{ij} - C^{Obs}\| \tag{5.12}$$

A representative image of a dynamic obstacle can be seen in Fig. 5.6 (Wolf and Burdick, 2008).

The total effect of all the obstacles in the workspace is given as

$$\mathbf{U_{Obs}} = \mathbf{U}_{Stat} + \sum_{n=0}^{N} \mathbf{U}^n_{Dyn} \tag{5.13}$$

Where N is the total number of obstacles observed. The extremely large values that are generated are truncated to a maximum viable value.

5.3.4 Modelling the Points of Interest

A Point of Interest (an inexhaustive list of what may be considered as a POI can be found in Helbing and Molnar (1995)) generates an attractive pull in the scene. With sufficient motivation, the self-driven particle can escape the influence of a POI. A POI is also a terminal point in the scene - the implication being that all exits in the scene are POIs. The potential of a POI situated at a cell defined by (x_{poi}, y_{poi}) is a Gaussian function centered at (x_{poi}, y_{poi}). $\beta_{poi}, \sigma_x, \sigma_y$ depend on the global importance of the specific Point of Interest.

$$\mathbf{U}^n_{POI} = \beta_{poi} \exp\left(-\left[\left(\frac{X - x_{poi}}{\sigma_x}\right)^2 + \left(\frac{Y - y_{poi}}{\sigma_y}\right)^2\right]\right) \tag{5.14}$$

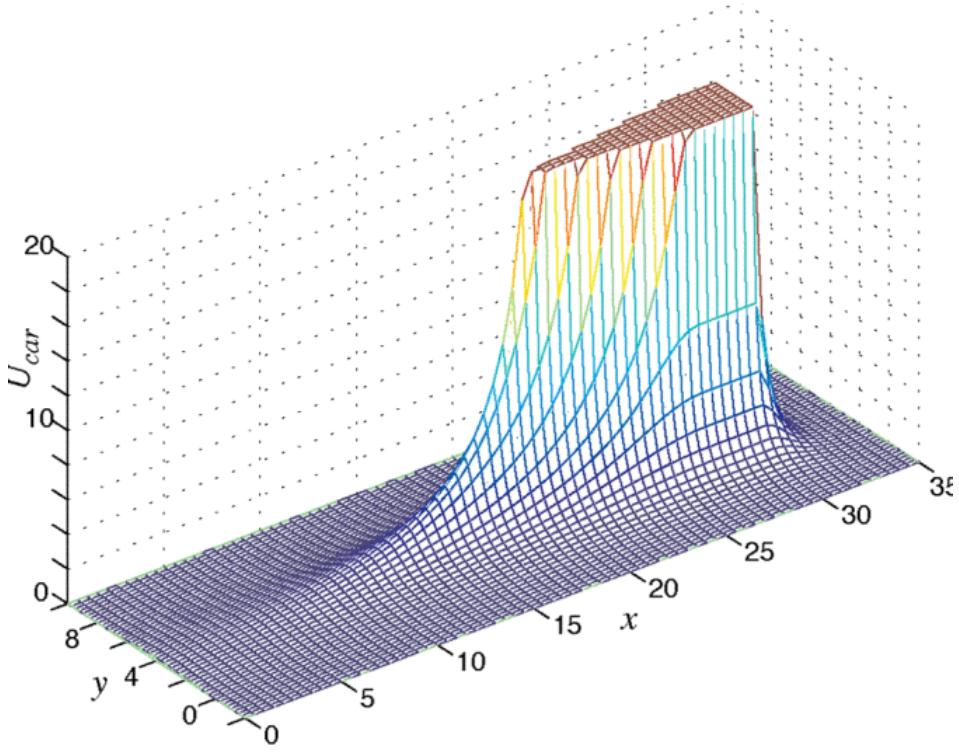


Fig. 5.6.: Potential cost of a dynamic obstacle. A triangular wedge is added in the direction of motion to compensate for future costs that can be incurred [Image credit: (Wolf and Burdick, 2008)].

Let $\mathcal{P}$ be a set of length N containing the positions $\langle x_n, y_n \rangle$ for each of the N POIs observed in the scene.

Subsequently,

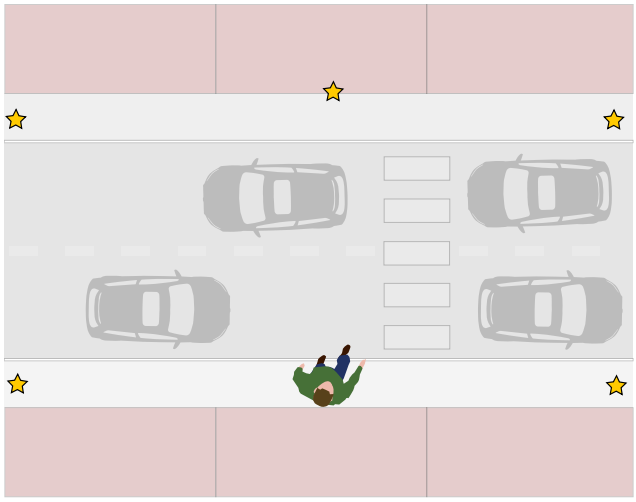
$$\mathbf{U}_{POI} = \sum_{n=0}^{N} \mathbf{U}_{poi}^{n} \tag{5.15}$$

$$J([\Psi = POI]) = \mathbf{U}_{POI} \tag{5.16}$$

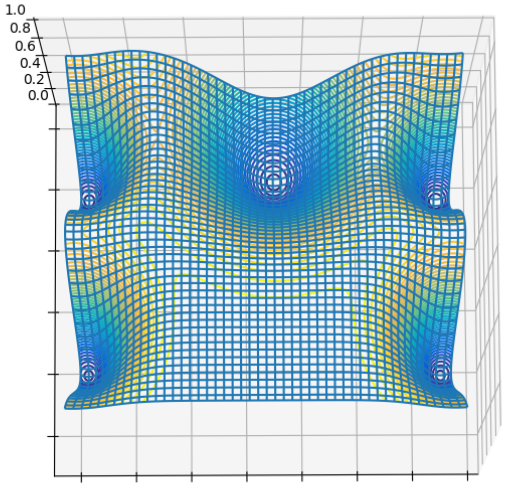
Fig. 5.7 illustrates the effect of the different POI on the pedestrian. A pedestrian wanting to cross the street is affected by the attractive nature of the POIs causing him to move towards them, thereby causing him to cross. As is observed, the existence of these POIs reduce the cost of the area where they are present.

5.3.5 Building Potential

When we consider a built, urban environment, we identify the effects the features present within the environment. Such a scenario demands the exclusion of buildings in so far as



(a) Highlighted "POI" semantic feature of the typical urban scene, depicted using golden stars.



(b) Potential "POI" Cost for the typical scene depicted above. The point POIs have an attractive area surrounding them illustrated by the lines directed into the troughs in the figure.

Fig. 5.7.: POI Potential for the typical urban scene depicted above.

nothing but obstacles, thus being impassable. On the other hand, entrances to buildings may be Points of Interest, explaining pedestrian motion towards them.

Theoretically, for modelling, a building can be considered a static obstacle such that $J([\Psi = Building]) = \mathbf{U}_B$ becomes:

$$\mathbf{U}_B = \begin{cases} u_{max} & \text{if } m_i = Building \\ 0 & \text{Elsewhere} \end{cases} \tag{5.17}$$

Once this cost is calculated, it is normalised. The effects of buildings in the potential cost can be seen in Fig. 5.8 for the typical urban scene as before. In the figure, the building semantic feature can be seen to be at the highest possible cost (1.0) making these areas impassable.

5.3.6 Total Potential Cost of the Environment

The total potential cost due to the different observed elements in the observed scene is the summation of the constituent potential costs.

Let $\Theta = [\theta_{Road}, \theta_{Edge}, \theta_{Obs}, \theta_{POI}, \theta_B]$ be the contributing weights of the different potential costs. Similarly, let $\mathbf{U} = [\mathbf{U}_{Road}, \mathbf{U}_{Edge}, \mathbf{U}_{Obs}, \mathbf{U}_{POI}, \mathbf{U}_B]$ be the vector of the constituting potential costs.

Then, the total potential cost of the entire observed scene can be defined in terms of Ψ can be defined as:

$$J(\Psi) = \mathbf{U}_{Total} = \Theta \cdot \mathbf{U}^\top \tag{5.18}$$

Since the different weights depend on factors like the width of the road L_{Road}, etc, which have already been included in the individual cost calculations, Θ can be replaced by a vector of 1's.

Thus, the total potential cost becomes:

$$\mathbf{U_{total}} = \mathbf{U}_{Road} + \mathbf{U}_{Edge} + \mathbf{U}_{Obs} + \mathbf{U}_{POI} \tag{5.19}$$

The illustrative resultant of the different potential costs for the typical scene referred to previously is seen in Fig. 5.8. This cost map representation can be intuitively understood as follows -

- The red areas within the generated map are considered impassable areas - as in the case of the cars and the buildings. Thus, for the pedestrian, those areas cannot be traversable while considering a crossing.
- The orange lines close to the cars are areas which still pose a danger to the pedestrian. These are dependent on the speed and heading of the dynamic obstacle. Any mental plan the pedestrian always takes into account these repulsions.
- The green areas on the map are those around which a pedestrian can easily plan to traverse.

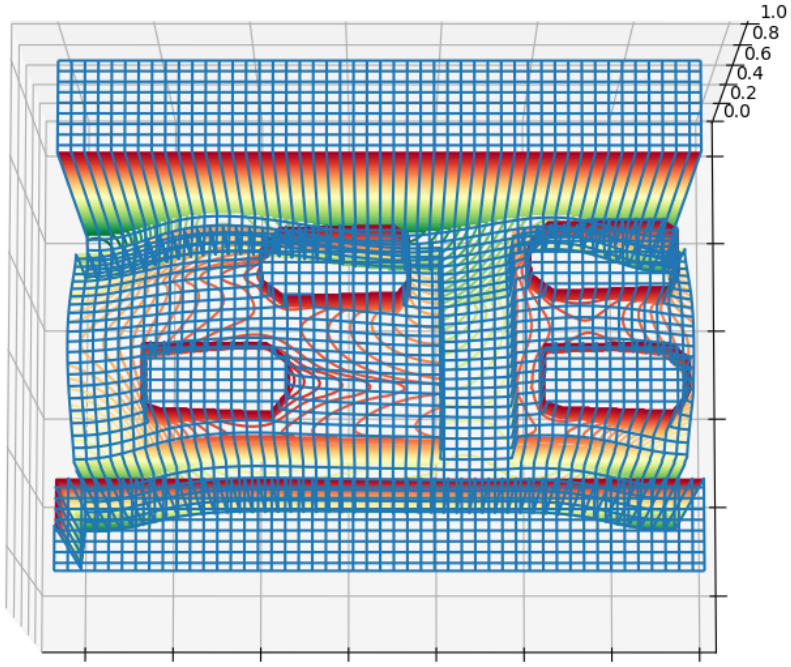


Fig. 5.8.: The resultant potential cost map of the illustrative typical scene of Fig. 5.2 depicting the summation of all the costs due to the identified features and the cost due to the dynamic obstacles. The Points of Interest's influences can be seen as depressions in the regions where they exist in Fig. 5.2.

- Blue areas are those which are still possible to travel by the pedestrian but with a higher awareness of caution.

Thus, from Fig. 5.8, while recognizing that it is a snapshot of the environment at a specific instant and that the real environment is dynamic, it is surmised that one of the safest ways for the pedestrian trying to cross from his current location towards a destination is through the crosswalk where the effects of the slow-moving obstacles is non-negligible yet not fatal. On the other hand, taking the quickest path requires him to negotiate between dangerous areas wherever allowed to arrive at his goal. Given this knowledge of pedestrian mental cost map, it becomes evident that knowing the complete state of the environment, it is highly probable that he is unwilling to undertake an illegal crossing. While the possibility of the action exists, such a trajectory can only occur in the "safe" areas of the scene that are not the crosswalk.

The algorithm for calculating the potential cost of an observed environment is given in Algorithm 5

5.4 Experimental Validation of the Proposed Model

To validate our proposed approach in modelling the environment, we compare it qualitatively against Inverse Optimal Control (IOC) modelling presented in Kitani et al. (2012). The referred work proposes a generative model for determining the cost function of a *static environment* skewed towards pedestrian preferences. They achieve this by preprocessing the observed environment through a classifier to identify various features present within and subsequently use tracked pedestrian trajectories to learn optimal weights for each of these features. This work leverages the IOC technique to learn the environmental cost map while using Optimal Control (OC) techniques over it to forecast a pedestrian's trajectory within this environment in a human-like mobility distribution between a starting and terminating point. Our work differs from that of Activity Forecasting (AF) model (Kitani et al., 2012) in the cost map generation, while utilising the same features in the environment. We qualitatively compare our proposed approach with that of AF in being able to forecast similar pedestrian distributions over time.

5.4.1 Dataset and Configuration

We use an extract of the VIRAT dataset (Oh et al., 2011) used in Kitani et al. (2012), to maintain fidelity in the qualitative comparison. This extract consists of different scenes in a parking lot comprising a building, stationary cars, grass, pavement and sidewalk. The pedestrians in each of the scenes are tracked using the Super-Pixel-Tracker (SPT) (S. Wang et al., 2011). The feature maps of the environment and pedestrian ground truths with tracking data along with its rectified images have been provided by Kitani et al.[2]

As our proposed approach requires as input a labelled image, we manually label the rectified (top-down) images provided in the dataset extract using the LabelMe tool (Wada, 2018) under the classes Pavement, Sidewalk, Grass, Building, POI and Obstacles. The labelled image with the underlying configuration of the environment is shown in Fig. 5.10. A comparative feature map for the environment can be seen in Fig. 5.9.

5.4.2 Qualitative Analysis

The first step in our comparison is the generation of cost maps via both approaches for the same environment. To generate the cost function depicted in Fig. 5.11a, we use the labelled image from Fig. 5.10 and parse it through the usage of Algorithm 5.

The cost map depicted in Fig. 5.11b is generated via Kitani et al. (2012) using the optimal weights of the features learnt on observing many pedestrian trajectories. Thus, a forecasted pedestrian distribution is expected to follow a "human-like" pattern, avoiding obvious obstacles such as cars, which can be seen in Fig. 5.12b. This figure depicts the result of the

[2] Activity Forecasting dataset/code: http://www.cs.cmu.edu/~kkitani/datasets/index.html

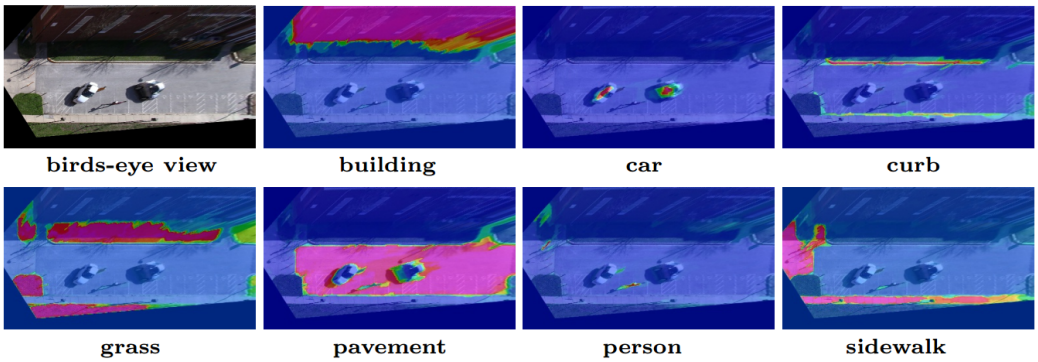


Fig. 5.9.: Feature Map of the VIRAT environment as an output of a classifier [Image Courtesy: Kitani et al. (2012)].

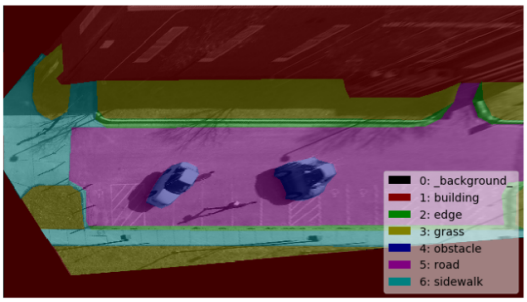
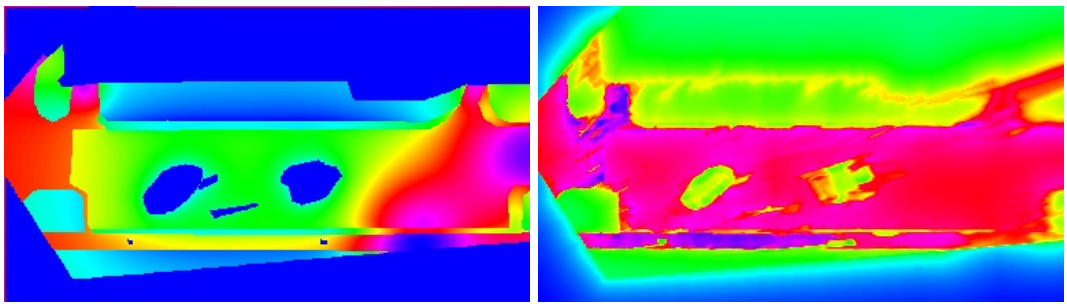


Fig. 5.10.: Rectified, Labelled Image from the VIRAT Dataset. The POI are presented with the small black dots.

AF approach of a higher reasoning of a pedestrian in the environment in traversing the area. It is seen that the forecasted distribution presents a scenario where the sidewalk is clearly preferred in reaching the goal while it is possible to take other trajectories traversing the pavement around the obstacles. Similarly, in replacing the cost map generated using the feature maps with one generated with our proposed method shown in Fig. 5.11a, the OC method arrives at a forecasted distribution depicted in Fig. 5.12a. We observe that qualitatively, the forecasted distribution imitates the original distribution as that of AF. The pedestrian prefers moving in the "safe" area of the sidewalk, avoiding higher cost areas. The acquired result resembles the AF method for human-like movement even for passing between the cars on the pavement.

Nevertheless, the generated cost function in AF (Fig. 5.11b) is dependent on witnessing multiple pedestrian trajectory demonstrations to find the optimal weights, which goes against the "Look-Once" principle for this modelling to be utilised on an autonomous car.

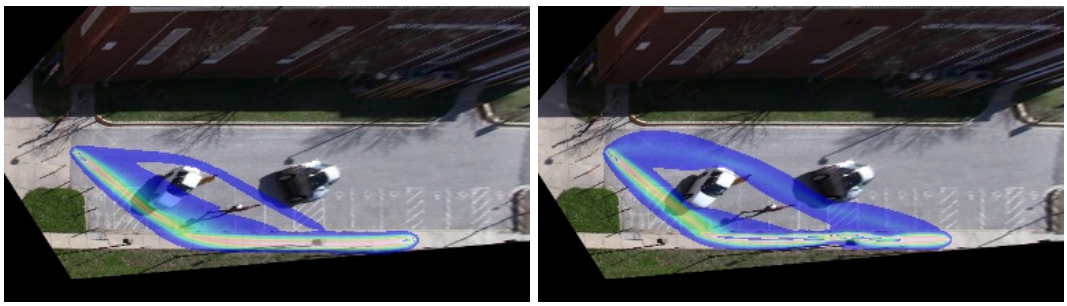
Consider the scene in Fig. 5.13 in the same environment as before but with a configuration in which the obstacles are placed differently. In this example, a pedestrian wants to move between the starting position in the environment and the terminating point.



(a) Cost Map generated using the labelled image presented in Fig. 5.10. The deep blue areas depict areas of high cost, i.e., that of unknown areas and obstacles while areas with hues of red to green depict areas of low to medium costs respectively. (Proposed)

(b) Cost Map generated with the feature map presented in Fig. 5.9. Blue areas in the figure depict unknown areas and thus areas of high costs. In contrast to the image on the left, Green depicts impassable areas with high cost, including obstacles. Areas with reddish hues are those that are conducive for pedestrian movement. (AF)

Fig. 5.11.: Comparison of Cost Maps on the basis of the methods of their generation.

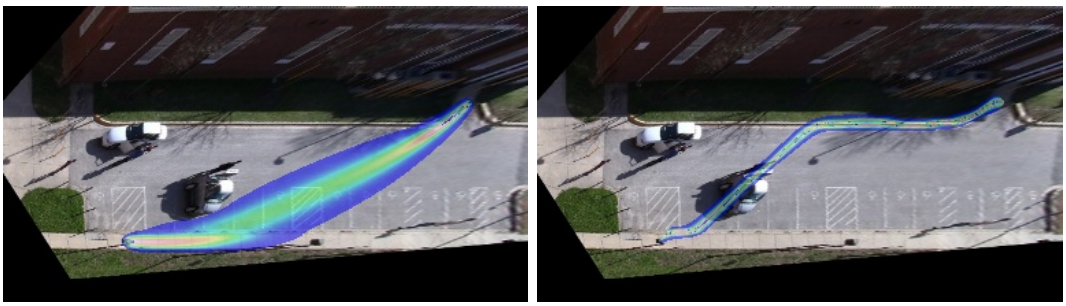


(a) Forecasted Pedestrian trajectory based on the cost map shown in Fig. 5.11a. (Proposed)

(b) Forecasted Pedestrian trajectory based on the cost map shown in Fig. 5.11b.(AF)

Fig. 5.12.: Forecasting of "human-like" pedestrian trajectories using Optimal Control based on generated cost maps.

The second qualitative test we perform is to evaluate whether our proposed method and the AF methods conform to the "Look-Once" principle. We generate the cost function using algorithm 5 for the configuration of the scene presented in Fig. 5.13. For the same scene, we generate the cost function using the feature map as described in the AF method with a single demonstration. We forecast the pedestrian distributions using both the cost maps and the results can be seen in Fig. 5.13a and Fig. 5.13b. The forecast in Fig. 5.13b devolves into a simple planner between the pedestrian's starting and termination points since the IOC cannot accrue enough information about the weights of the different features in the scene on a single demonstration. This leads to a distribution passing through an obstacle, where it clearly should not. On the other hand, Fig. 5.13a utilises inherent knowledge of the environment based on sociological cues providing a distribution that cleanly avoids the



(a) Forecasted Pedestrian trajectory with cost map generated using the proposed method, conforming to the "Look-Once" principle.

(b) Forecasted Pedestrian trajectory where the cost map for the environment is generated with only one demonstration with IOC which fails in forecasting a humanlike distribution.

Fig. 5.13.: "Look-Once" comparison of pedestrian forecasting between the proposed method and the IOC method.

obstacles while maintaining a "human-like" profile on the generated distribution thereby demonstrating conformity of our proposed method to the "Look-Once" principle.

5.5 Conclusion

In this chapter we proposed a novel method to model the environment utilising the different features in an observed scene wherein the sociological effects of each of these features factored into finding a quantifiable function affecting pedestrian motion in that self same scene. We introduced these sociological concepts, especially the principles of *Natural Vision* and *Natural Movement* which we used as a basis in understanding human behaviour. Further, we expounded on the idea of a Point of Interest modifying pedestrian behaviour in built, urban environments. We validated our approach by qualitatively comparing the "human-like" motion generation using the AF approach from Kitani et al. (2012) which exploits a cost function learnt on human demonstrations. We perform this by replacing the demonstration based cost map with our sociological based cost map and generating pedestrian distributions between two points in a static environment. Second, we validated that our approach was able to conform to the "Look-Once" principle in Section 5.4 by showing the generation of human-like pedestrian distributions where the AF approach failed.

In contrast to existing approaches that model the environment (Ziebart, Ratliff, et al., 2009; Kitani et al., 2012) on demonstrated trajectories in time invariant environments, our approach can encompass time varying environments which are frequently observed in urban environments. Another advantage of our approach over existing ones is the reduction of complexity. Inverse Optimal Control methods have a complexity non-linear in the demonstrations provided for learning (Ziebart, 2010) while the proposed approach has no such dependencies but is linear in the number of observed features. The proposed model also

differs from existing literature in the choice of destinations, where some such as Deo and Trivedi (2017) cluster sources and destinations of demonstrated trajectories to designate goals or consider all positions in the observed environment as possible goals (Kitani et al., 2012), we assume, with sociological backing, that there are only a few viable destinations in any observed scene.

Over the course of this chapter, we have spoken about structured environments and the influences they have on pedestrian behaviour. Unstructured environments, it must be noted, do not affect behaviour similarly. Thus, the proposed method fails to perform well within bounds of such an environment. Another shortcoming of this approach is its reliance on strong perception systems. One such is the requirement of needing a well identified feature set for the scene. An unintended consequence of this is the need to identify Points of Interest in this scene as well. Both of these related shortcomings can be resolved by using, for example, a detailed map of the region the observer is present in. This map could contain geometric and semantic information in the vicinity of the autonomous vehicle. Consequently, the problem reduces to one of localising itself using proprioceptive and exteroceptive sensors, the solution to which is out of the scope of this thesis.

In the following chapter, we propose an extension to a graphical model that leverages this cost map to predict pedestrian positions, even before any observations are made. Moreover, the proposed framework utilises the presented cost map generation approach as the first part of a pipeline that is meant increase the effectiveness of pedestrian prediction from an autonomous car.

Algorithm 5 Total potential

Input
Road Geometry L_{Road}
Feature List Ψ
POI Information List $\mathcal{P}$
Environment Image X
Obstacle List O
Tracker output

Output
$\mathbf{U}_{Total}$

procedure PotentialCostMap(Ψ, $\mathcal{P}$, L_{Road}, X)
  X $\leftarrow$ discretize(X)
  for $\psi \in \Psi$ **do**
    if $\psi \equiv Road$ or $Edge$ **then**
      $\mathbf{U}_{\psi} \leftarrow \mathbf{U}_{\psi} + \mathrm{J}(\mathrm{X}, [\Psi = \psi])$ ▷ Using Eqn. 5.4 and Eqn. 5.7
    else if $\psi \equiv Building$ **then**
      $\mathbf{U}_{\psi} \leftarrow \mathbf{U}_{\psi} + u_{max}$
    end if
  end for
  $\mathbf{U}_{\psi} \leftarrow$ normalise$(\mathbf{U}_{\psi})$
  for all $p \in \mathcal{P}$ **do**
    $\mathbf{U}_{POI} \leftarrow \mathbf{U}_{POI} + \mathrm{J}(\mathrm{X}, p)$ ▷ Using Eqn. 5.14
  end for
  $\mathbf{U}_{POI} \leftarrow$ normalise$(\mathbf{U}_{POI})$
  for $o \in O$ **do**
    if $type(o) = Static$ **then**
      $\mathbf{U}_{Stat} \leftarrow \mathbf{U}_{Stat} + u_{max}$ ▷ $u_{max} = 1$
    else if $type(o) = Dynamic$ **then**
      $v \leftarrow$ getVelocity(o)
      $obj \leftarrow$ extendObstacle(o, v)
      $\mathbf{U}_{Dyn} \leftarrow \mathbf{U}_{Dyn} + \mathrm{J}(obj)$ ▷ Using Eqn. 5.11
    end if
  end for
  $\mathbf{U}_{Obs} \leftarrow \mathbf{U}_{Stat} + \mathbf{U}_{Dyn}$
  $\mathbf{U}_{Obs} \leftarrow$ normalise$(\mathbf{U}_{Obs})$
  $\mathbf{U}_{Total} \leftarrow \mathbf{U}_{\psi} + \mathbf{U}_{POI} + \mathbf{U}_{Obs}$
  $\mathbf{U}_{Total} \leftarrow$ normalise$(\mathbf{U}_{Total})$
end procedure

6 Learning and Prediction of Motion

> *The best way to predict the future is to create it.*
>
> **— Dr. Forrest C.Shaklee**

In chapter 5, we presented a method of modelling the observed environment using Gibson's theory of *Natural Vision* (Gibson, 1979) and potential fields (Vasishta et al., 2017). In sections 6.2 and 6.3, we present our proposal called the Semantically-Aware ITM (S-ITM) that utilises the real environment to build and update a *Prior Topology*, following which we discuss our modifications to the conventional GHMM model for learning the structure and parameters based on pedestrian observations and Prior Knowledge. In section 6.4, we discuss inference and prediction for pedestrians in static and dynamic environments. We conclude the chapter with discussions on the presented model.

6.1 Partial Trajectory Utilisation

In Section 4.4, we discussed some of the shortcomings of the GHMM method. One of the drawbacks mentioned in this discussion was the need for full trajectories so as to generate the augmented observation ω (Eqn. 4.1). In scenarios where the full trajectories are unavailable, or is infeasible, such as the online observation and tracking of pedestrians from a moving platform, this method is rendered potentially unusable. This unsuitability stems from the fact that most real world tracking systems lose track of their objects, for example, due to occlusions. Therefore, the terminating observation of such sequences are not the true destinations. In fact, it could be possible that the tracked object is lost momentarily and re-tracked but as a separate object whose trajectory could then terminate at the true destination. Another shortcoming as put forth in the same section is the way in which the topological map is learnt. To reiterate, the nodes $n \in \mathcal{U}$ are all associated with Gaussians whose mean μ is changing but the co-variance matrix Σ remains the same. Such an assumption implies that the directionality of pedestrian motion is irrelevant in the process of estimating a pedestrian's future position. This has been proven to be untrue as in the cases of Turner and Penn (2002).

A simple solution is to utilise partial trajectories instead of full trajectories for the learning of the topologies and the parameters of the GHMM. In this case, the problem then devolves to one of identifying pedestrian goals, i.e., to populate the set Φ. Once this set has been populated, these goals[1] can then be used to generate augmented observation (restated here for convenience):

$$\omega = (O_t^k, \phi), \{\forall\phi | \phi \in \Phi\} \tag{6.1}$$

In Eqn. 6.1, O_t^k denotes the t^{th} observation in the k^{th} sequence. In turn, they are used to initiate the state S (Eqn. 4.3) where n is a node on the topological map.

This is achieved by quantifying the intuition that over a sustained period of time in a given environment, pedestrians tend to appear or disappear in specific regions. Another cue in recognising transitory destinations is to observe the time spent by people in different regions of this environment (Luber et al., 2011). Ostensibly, certain areas of the environment are more conducive as destinations for pedestrians. For example, reconsider the environment described in Fig. 4.2. Transforming this environment into a corridor and imagining a desk in place of an obstacle, it is not inconceivable that some pedestrians stop for a while at this desk before continuing to their destinations which could be any of the three visible exits. Continuing this example, if we were to imagine a vending machine in any corner of the corridor, it is once again possible to think that some pedestrians stop there momentarily longer than in other areas. Such behaviour can also be seen in other closed environment studies such as Z. Wang, Ambrus, et al. (2014). The lessons learnt from these studies can be generalised to apply to an outdoor pedestrian environment. Instead of, say, a vending machine, a pedestrian in an outdoor environment could just wait longer at a traffic light to cross the street. Areas from where they might appear or disappear, for a tracker, could then be entrances and exits of commercial establishments or houses.

One such method to automatically discover pedestrian goals based on their waiting times was proposed by Pérez-Hurtado et al. (2015). Their work extends the GHMM method presented in Chapter 4 in the following ways:

- Automatic discovery of goals based on observed pedestrian trajectories
- Modification of the GHMM method to utilise partial trajectories
- Extension of the learning phase to also learn a better sensor model $P(O_t \mid S_t)$

As a note on the nomenclature, we define partial trajectories here as those trajectories whose origin and destinations are not known *a priori*. Instead, such trajectories are considered a part of a longer *full trajectory* with a well defined start and end points. For all

[1] As previously, we use the term goals and destinations interchangably. Goals and by extension, destinations, are those regions in the observable environment where pedestrians can either begin or terminate their trajectory.

partial trajectories, the point at which they are identified by a local tracker i.e., assigned a tracking number k is considered their origin, regardless of the fact that the associated pedestrian could have been assigned another tracking number before being occluded or lost to the tracker. Their destinations, as a part of the termination of the partial trajectory, is considered irrelevant. As previously, these partial trajectories are mathematically treated the same as a full trajectory with their sequence denoted by $O^k_{1:T}$ - the k^{th} partial trajectory.

Automatic Discovery of Goals The discovery of goals is a result of a modification in the ITM method used in the original GHMM method. When an observation ω, $\omega \in O^k_{1:T}$, it modifies the topological space $\mathcal{E}$ differently. Each node n is associated with a Gaussian $G(n; \mu, \Sigma)$ as previously but also with tuple of counters $\langle n_n, n_{in}, n_{out}, n_s \rangle$ each of them keeping count on the number of people traversing the node, people entering, exiting and the mean time people spend at this node n. With this information, an estimation can be made on whether a node is an entry/exit point with the simple assumption that given a threshold value, if over time, there are many pedestrians appearing or disappearing there, it is possible to designate those nodes as destination points, thereby populating Ψ. A second category of goals which we previously discussed is found in areas where pedestrians wait longer than usual. This information can be obtained using the counter n_s wherein if the counter, over time, is longer than a threshold value, the corresponding node can be designated a destination too.

Partial Trajectory Utilisation An important contribution of Pérez-Hurtado et al. (2015) is the modification of the GHMM method to utilise partial trajectories for learning the parameters of the underlying HMM model λ. This is achieved by altering the partial trajectory so as to be equivalent to a full trajectory, thereby freeing up the space for the utilisation of incremental algorithms to estimate the parameters like the Incremental Baum Welch method. This alteration involves, first, the discovery of possible goals in the environment. Secondly, it does away with the need for an augmented observation ω, instead relying on the instantaneous observation O_t. The intentionality of pedestrian movement is dissociated from the observation and instead, becomes a function of the topology and the graphical model. Once the goals are discovered, consequent steps follow the updating of the structure as set out in subsection 4.2.2. This procedure is run on every observation of every partial trajectory since, on every run of the ITM, there can be new nodes generated or removed and subsequently, a modification of the edges too - either by addition, subtraction or reordering. Regardless, this modification allows the partial trajectory to update the structure of the GHMM.

Sensor Model Update The final contribution of Pérez-Hurtado et al. (2015) is the Sensor Model Update. As mentioned in the previous paragraph, the intentionality of pedestrian

motion relies now not on the observation itself but on the different spaces defining the GHMM, specifically the topological space $\mathcal{E}$ and the Model State Space $\mathcal{S}$. In dissociating the intentionality, a simple assumption is made - pedestrian motion generally points in the direction of a destination. Such an information, which can be inferred from the observation $O_t = (\mathbf{x}_t, \mathbf{v}_t)$, can be encoded within the topological map itself. All nodes in the topological map are associated with their own Gaussian $G(n)$, defined by its mean μ_n and its covariance Σ_n, a change from the classical implementation wherein the covariance matrix Σ remained unchanging for all nodes of the topology. On every new observation $O_t = (x, y, v_x, v_y)$, if it is associated with the node n, then the parameters of the Gaussian distribution associated with this node is updated with both the mean and covariance changing as the average and the standard deviation, respectively, of the incoming observation and the existing parameters.

The Sensor Model, or the Observation Probability, for any state S_t is defined as $P(O_t \mid S_t)$. Our observation at time t is composed of the positional and velocity component $(\mathbf{x}, \mathbf{v})$ while the state is composed as a tuple of the corresponding node n and a goal ϕ. Thus, the observational probability can be written as:

$$P(O_t \mid S_t) \doteq P\left((x, y, v_x, v_y) \mid (n, \phi)\right) \tag{6.2}$$

Using the probabilistic laws defined in Chapter 3, Eqn. 6.2 can be decomposed as follows.

$$P\left((x, y, v_x, v_y) \mid (n, \phi)\right) = P\left(v_x, v_y \mid x, y, n, \phi\right) P\left(x, y \mid n, \phi\right) \tag{6.3}$$

The second component is the Gaussian distribution associated with the node n as described previously. Thus, each treated observation aids in learning a better Observation Probability model.

While there are many advantages to the described method, there are also some non-trivial shortcomings.

Drawbacks

1. Automatic Goal Discovery is a lifetime learning process wherein goals are discovered based on observations updating the parameters of nodes $n \in \mathfrak{U}$, $(n_n, n_{in}, n_{out}, n_s)$. Longer the number of observation sequences, greater the chances of representation along all nodes and thus, the "discovery" of more goals and the removal of previously discovered goals. As a consequence, states $S = (n, \phi)$ are frequently created and destroyed, leading to a ballooning of complexity of the model.

2. The Improved Instantaneous Topological Update manages to delegate the intentionality of the pedestrian trajectory to the topological map and the state space. Yet, this approach does not increase the semantic information available for better inference. Directionality and lingering are considered as cues for points of interest in the envi-

ronment. Such a straightforward correlation between the two becomes meaningless with no contextual information. For example, pedestrians might linger at a bottleneck in their path or a temporary obstacle might cause an occlusion leading to people "entering" from an associated node. Such areas have no semantic bearing grounded in the real environment and cannot be identified as a door or other points of interest.

3. Following the original implementation of the GHMM, this extension cannot infer pedestrian positions where there have been no previous partial trajectories observed.

4. The improved ITM algorithm proposed in Pérez-Hurtado et al. (2015) cannot guarantee the convergence of the topological map.

While the extension proposed in Pérez-Hurtado et al. (2015) solves some of the issues highlighted in the original GHMM implementation, as described above, it inherits much of the problems discussed in Section 4.4.

Any method proposed to tackle these problems needs to meet the following **Requirements**:

1. Goals in an observed environment need to be fixed and unchanging.

2. A generative model of the topological map needs to *a priori* account for all the areas of the observed environment

3. A generated topology needs to be contextually aware of the semantics of the environment that it is generated in, such as awareness of obstacles and their impermeability.

4. The underlying graphical model must be able to anticipate pedestrian behaviour and, if inaccurate, converge to such a model as to minimise these errors.

6.2 Semantically Aware Topological Mapping

Topology Representing Networks were previously discussed in Section 4.2, where a form of vector quantisation was used to map an incoming input to create creases in a manifold i.e., the observed environment, and thus "discover" the underlying topology of this environment. A Topological Map G, discretizing the Space $\mathcal{E}$, should not only map observations to this Space but must also take into account the constraint set out in Requirement 3. Thus, the function $f(\omega)$, mapping the Observation Space $\boldsymbol{\Omega}$ to $\mathcal{E}$ needs to be redefined. This redefined function $h(\omega)$ can then be thought of as a composition of $f(\cdot)$ and another function $g(\cdot)$, representing the constraints.

This section describes one of the main contributions of the thesis - the Semantically-Aware ITM (S-ITM) algorithm. Consider an encumbered environment with obstacles. In such an environment, pedestrians, obviously, cannot pass through the obstacles and should

find their way around them. Any algorithm proffering to solve the problem of representing trajectories should be cognisant of the semantics of the observed environment and its affordances[2]. One approach to ensure that these requirements are met is to generate *a priori* a topology representing all possible pedestrian paths in the affordable area.

Since it is impossible to represent all possible pedestrian paths in any given environment, we relax the requirements such that this *a priori* topology, which we call the ***Prior Topological Map (PTM)***, should:

1. **Represent:** typical trajectory prototypes,

2. **Associate:** trajectory prototypes with the semantics of the observed environment so as to capture all possible entrances, exits and points of interest contained within,

3. **Learn:** new trajectory representations based on new observation sequences.

At this juncture, it becomes evident that the function $h(\omega)$ is a multivariate function composition in $f(\cdot)$ defined as:

$$f\ |_{\omega_0=g} \doteq f(g(\varnothing), \omega_1, \omega_2, \ldots, \omega_n) \tag{6.4}$$

where $\omega_0, \omega_1, \omega_2, \ldots, \omega_n \in \Omega$ and $\omega_0 = \varnothing$ is a special case of the observation space.

Eqn. 6.4 defines the function f as composed of the function $g(\varnothing)$ which generated a Prior Topological Map that does not require any observation. The topology generated by this function is acted upon sequentially by $f(\cdot)$ on the various observations $\omega \in \Omega$.

The proposed Semantically-Aware ITM (S-ITM) algorithm builds on this discussion and can be divided into two phases, reflected in the organisation of this section. The first phase is the "Constitution" phase, where the Prior Topological Map (PTM) is instantiated within the observed environment with the corresponding parameters and associated attributes of the nodes of the topology. The second phase, called the "Update" phase, corresponds to the learning of the atypical behaviours in the environment which, as a consequence, updates the topology.

6.2.1 The Prior Topological Map

Consequent to the "Constitution" phase of the S-ITM, a Prior Topological Map must be generated corresponding to the semantics of the environment. The sub-problem in the "Constitution" phase becomes that of selecting valid nodes and connecting them via edges.

[2]In a pedestrian environment, the degree to which a pedestrian can walk in an area is the *affordance* (Gibson, 1977) of this area.

Configuring the Prior Topological Map

In the search of a solution to the subproblem described previously, the following assumptions are made:

1. There exists a function J that maps the affordances of the different areas of the observable space (Ω) to quantifiable costs for the different features Ψ.
2. There are only a finite number of destinations in the observable space and that all destinations are known.
3. Nodes cannot exist within regions encompassing semantic feature labels denoting obstacles.
4. Nodes must be instantiated such that all affordable regions of the observable space are reachable.

In the quest for $g(\varnothing)$, conforming to Requirement 1, all viable, identified destinations ϕ are included in the set $\mathcal{V}$ such that:

$$\mathcal{V} = \{\forall \phi : \phi \in \Phi\} \tag{6.5}$$

with ϕ, in turn, being a vector in $\mathbb{R}^2$.

In satisfying Requirement 2, a population of p observations are generated from the positional subspace in Ω. To decide which of these observations are valid to populate the set $\mathcal{W}$ – containing nodes representing the affordable areas – we propose a two step process of passing a trial and sampling this reduced set.

The first trial is used to determine whether the observation p is contained within an area which is designated as an obstacle.

Given the cost function J such that $\mathbf{U_T} = J(\cdot)$, then the observation p can be a part of the *valid points set* K such that,

$$K = \{p \mid \mathbf{U_T}[p] < 1\} \tag{6.6}$$

Eqn. 6.6 results in a set containing valid observations from which the nodes of the topological map are chosen. The population of K is sampled using the Poisson sampling technique. Each element in the population has a distinct probability in getting chosen depending on the probabilities $q_1, q_2, q_3, \ldots, q_n$ for the n valid semantic labels of Ψ[3]. A second sampling is performed within this reduced set to check for the insertion criterion based on τ, the insertion distance. The nodes that pass this test compose $\mathcal{W}$. Thus we can define the set of valid prior nodes $\mathcal{U}$ as:

[3]This is a subset of Ψ since semantic labels like "car", "obstacle", "building" etc. are disregarded. For simplicity, we continue referring to this subset as Ψ.

$$\mathcal{U} = \mathcal{V} \cup \mathcal{W} \tag{6.7}$$

The Prior Topological Map itself is generated by the utilisation of the Delaunay Triangulation. In Vector Quantisation, of the p observations that are valid, these observations leading to nodes act as the centroids of the *Voronoi Tessellation* of the observed space. The Delaunay edges, which are the dual of these Voronoi regions, connect pairs of adjacent regions to each other (Fig. 6.1). If the edge passes through an obstacle, then such an edge is discarded.

These pairwise connections between two nodes form the links $\mathcal{L} = \{l_{i,j} : \forall i, j \in \mathcal{U}, j \in \aleph(i)\}$[4] of the Prior Topological Map. One condition to note is that the edges cannot pass through obviously impossible areas, for example, through an obstacle i.e., where $\mathbf{U_T}[p] = 1$

The Prior Topological Map is the solution of the function $g(\varnothing)$ and is defined by a tuple of the nodes and links, $\mathcal{E} = \langle \mathcal{U}, \mathcal{L} \rangle$.

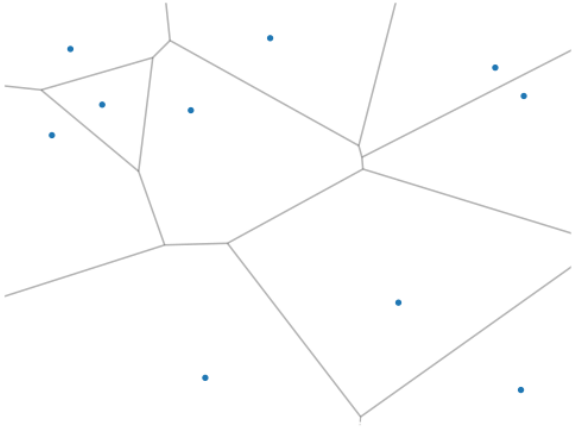


Fig. 6.1.: Voronoi regions of the observed space. The centroids of these regions are candidates for nodes of the PTM. Centroids of the observed area that exist within forbidden areas in the real world are automatically discarded.

A test on all elements of $\mathcal{L}$ are performed to ascertain that they do not pass through impassable areas. The links that fail this test are removed.

Constituting the Gaussian Probabilities

Each node thus chosen to form the PTM are associated with a Gaussian distribution to act as one of the constituents of the observation probabilities. Thus, around each node of the PTM, a uniform Gaussian distribution is maintained. This Gaussian is centered on the position of the node with the covariance uniformly placed over all dimensions with a

[4] $\aleph(n)$ is the Adjacency function of node n, returning a set of all centroids of the Voronoi regions adjacent to to the Voronoi region defined by n

value corresponding to the insertion distance τ. For the n^{th} node located in 2D space, this distribution is composed as in Eqn. 6.8:

$$\mathbf{G}(\mathrm{n}; \mu_\mathrm{n}, \Sigma_\mathrm{n}) = \mathbf{G}\left(\mathrm{n}; (\mathrm{x_n}, \mathrm{y_n}), \begin{bmatrix} \sigma_{xn}^2 & = \tau^2 & \sigma_{xn}\sigma_{yn} & = 0 \\ \sigma_{xn}\sigma_{yn} & = 0 & \sigma_{yn}^2 & = \tau^2 \end{bmatrix}\right) \tag{6.8}$$

In contrast to the original implementation of the GHMM method (Vasquez et al., 2009) and following Pérez-Hurtado et al. (2015), the covariance of each of these Gaussians are deformable with Σ_n itself being one of the parameters to be updated in the update phase.

6.2.2 Updating the Topology

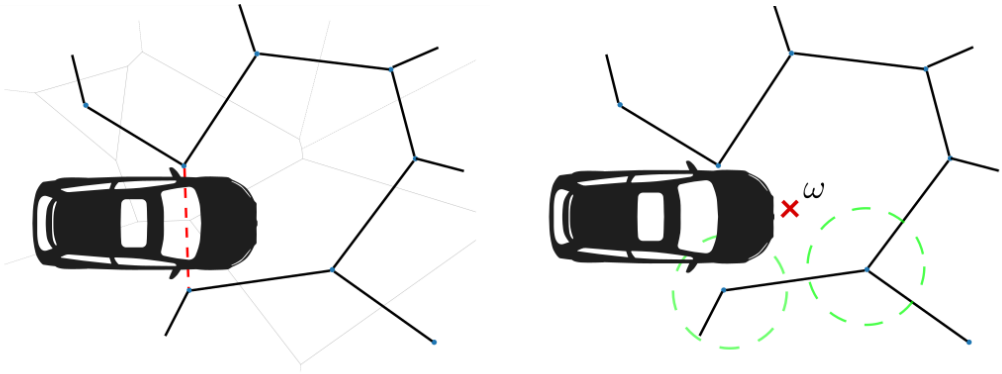
The Prior Topological Map is, primarily, a Self Organising Map (also called a Kohonen Network)(Kohonen, 1982). With regard to the relaxed requirements, only the representation and association of typical trajectories is insufficient. The model must also learn to represent trajectories that have not been captured with the prior topological map. Therefore, the model needs to be updated on new observations as they arrive.

It may be observed that the number of nodes of the PTM depend on the size of the observation space. Larger this space, higher the number of nodes selected from K, thus rendering an update process according to the original process unwieldy. Instead, we update the topology using Semantically-Aware ITM (S-ITM) where the update process is not only aware of the size of the topological map but also the underlying physical context of the environment.

Consider a part of the topological map, itself derived from the Voronoi tessellation of before, as shown in Fig. 6.2a. A part of an obstacle (hatched area) juts out into the environment where the PTM should exist. Also consider that there are observations sensed around this obstacle in a manner as to satisfy the insertion criteria. Within the confines of the existing implementation of the ITM (Vasquez et al., 2009; Pérez-Hurtado et al., 2015), an edge is created in the topological map signifying a transition between these two regions of the environment (denoted by the red dashed line in the figure) . On the contrary, the proposed S-ITM is aware of the existence of this obstacle and its inherent properties *i.e,* that these regions are impassable from each other. Thus, the creation of this pseudo-transition on the topological map does not occur.

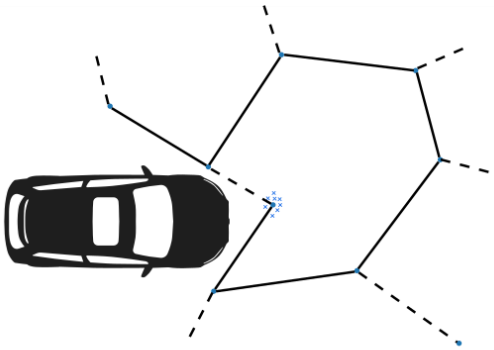
Fig. 6.2b and Fig. 6.2c illustrate the updating of the topological map via the S-ITM. An observation ω is perceived in the vicinity of, yet non-associable with, existing nodes *i.e,* it satisfies the insertion criteria. At this juncture, the S-ITM adds a node at this observation and associates it with this new node and the corresponding edges are added. This new node can then be associated with incoming observations that are perceived in its vicinity.

This topology can subsequently, under appropriate observations, can create an edge as depicted in Fig. 6.2c between this newly created node and the node on the other side of the obstacle. Such a path between the two regions separated by the obstacle is superior to



(a) The hatched part of the image signifies the existence of an obstacle. The centroid of the Voronoi region depicted in red is discarded since in the real environment, there cannot exist a node within an obstacle. Other centroids that conform to the insertion criterion are chosen as nodes and connected via Delaunay triangulations to create a topological map. Similarly, the edge that would've passed through the obstacle is removed since there cannot be a link between the two regions through the obstacle.

(b) An observation is senses which lies outside the areas associated with existing nodes, depicted by the dashed green circle depicting the insertion criteria. A node can now be created at this position.



(c) A new node is created and links join the areas on either side of the obstacle via the newly created node leading to a more human-like motion pattern.

Fig. 6.2.: Topological Map update using the Semantically-Aware ITM algorithm

the edge originally discarded in Fig. 6.2a since it conforms to a more natural, human-like trajectory.

A significant implementation bottleneck in this process is related to the size of the topological map itself. As the size of $\mathcal{U}$ increases, the complexity of node matching (algo. 4) between the incoming observation and the topological map to find the two closest nodes

to it increases linearly with the number of nodes. In large spaces, this becomes unviable rapidly. Instead, by initialising the nodes of the PTM in a K-d tree, the observation can be compared against this tree thereby reducing the matching complexity to an average of $\mathcal{O}(\log N)$.

Updating the Gaussian

This step of the S-ITM process corresponds to the Gaussian Adaptation phase of algorithm 4. Simultaneously, during the update of the topology, there is also an adaptation phase of the Gaussian around the node. This could take one of two forms - in the first form, if the node has been newly created, then a Gaussian is initiated around this node. The second form corresponds to that action where there is an already existing node which can accommodate the incoming observation. If this condition evaluates to true, the Gaussian around this node is adapted to better explain this new information. This adaptation, in essence, is the averaging of the means of the Gaussians with incoming observations and subsequently updating the standard deviation of the pre-existing Gaussians (Pérez-Hurtado et al., 2015).

6.3 Learning with Prior Knowledge

With the Prior Topological Map defining typical trajectories in an environment and the consequent topological update allowing for the deformation of this topology to represent new trajectories, it becomes imperative to initialise the underlying structure of the HMM as well as to learn pedestrian behaviour at its different states. In conforming to **Requirement 4**, not only must the underlying graphical model be able to anticipate pedestrian behaviour *a priori*, it must also learn diverging behaviour based on new observations. In this section, we describe the use of **Prior Knowledge**, i.e., using data about the environment to initialise and update the graphical model. As a result, this prior knowledge is dependent on the function J and the resulting $\mathbf{U_T} = J(\cdot)$.

6.3.1 Moving from Topologies to Graphical Models

The topologies generated in Section 6.2 need to be converted into graphical models to perform Bayesian Inference. Once more, in the previous section, we had assumed that all the destinations in the observed scene are known. Contrary to the process presented in Vasquez et al. (2009), which designates destinations based on the final observation in the trajectory or as in Pérez-Hurtado et al. (2015) where the destinations in the scene are learnt, these areas in the proposed approach are unchanging. Such an assumption allows for some leeway in reducing the complexity of the graphical model which is to be created.

Similar to the S-ITM method, the structure learning can be divided once more into two phases. At the end of the "Constitution" phase of the S-ITM, the following steps are executed to learn the structure of the underlying graphical model corresponding to the topological map $\mathcal{E}$:

1. Gather all possible goals $\phi \in \Phi$
2. for all nodes $n \in \mathcal{U}$, create all possible states (n, ϕ)
3. For each such created state, assign an appropriate prior value derived from $\mathbf{U_T}$
4. For every link in $\mathcal{L}$ with nodes n_i and n_j, create a transition between all the corresponding states for each goal.
5. Initialise each transition with appropriate transition values.

On every update of the S-ITM, with the arrival of a new $\mathcal{E}$:

1. For every removed node in the new $\mathcal{E}$, remove all the corresponding states from the graphical model and remove transitions to and from these states.
2. For every added node, create new states as previously with the same procedure for assigning priors and transitions

The assigning of the appropriate values for priors and transitions are described next. The values of the priors and transitions will be normalised at the end of this step.

6.3.2 Initial Probabilistic Model

The JPD of the underlying HMM created with the Prior Topological Map follows that of the GHMM as presented in subsection 4.3.1. For convenience, we restate it here.

$$P(S_{t-1}\ S_t\ O_t) = \underbrace{P(S_{t-1})}_{\text{state prior}}\ \underbrace{P(S_t \mid S_{t-1})}_{\substack{\text{transition}\\ \text{probability}}}\ \underbrace{P(O_t \mid S_t)}_{\substack{\text{observation}\\ \text{probability}}} \tag{6.9}$$

These three different components of the JPD are initialised differently as follows.

- ***State Prior Count:*** The state prior count $\boldsymbol{\pi}$ initiates the model λ. In a traditional model applied to our problem, this parameter is instantiated with a normalised count of pedestrians passing through the state. Since, in the GHMM, such a state is hidden, this count cannot be performed. On the other hand, given that the function J maps the affordances to real values, an analogue to this count can be obtained. Noting that the state S_t is composed of a node $n \in \mathcal{U}$ (exp. 6.7) whose position in Ω is two dimensional and with the definition that the affordance of an area says how likely it

is for a pedestrian to pass through it, we can define the state prior count for the i^{th} state, which is then normalised, as follows.

$$\pi_{\mathbf{0}}^{\mathbf{i}} \quad = \quad 1 - \mathbf{U_T}[n] \tag{6.10}$$

- ***Transition Probability:*** The transition probability between two states i and j, represented as $a_{ij} = P([S_t = i] \mid [S_{t-1} = j])$ in terms of our problem, is the probability of a pedestrian transitioning from one state to another over time. Transitions based on affordances[5] can happen in three forms -

  1. Transitions from a higher affordance to a lower affordance - For example, a pedestrian moving from an easily traversable area like a sidewalk to a more complicated area like the road.
  2. Transitions from a lower affordance to a higher affordance. Continuing with the same example, the pedestrian moving towards the sidewalk from the road.
  3. Transitions towards the same affordances as a pedestrian continuing to walk along the sidewalk.

  Studies (Brosseau et al., 2013; Thompson et al., 2013) have shown that pedestrians tend to move from areas of lower affordances to higher affordances or prefer to remain in areas of high affordances as compared to the contrary, so as to remain safe. We model the transition probability on this behaviour and make certain assumptions -

  – Transitions between areas of lower affordances to higher ones are to be rewarded.
  – Transitions between similar areas of affordances are neither rewarded nor discouraged.
  – Transitions from areas of higher affordance to lower ones are discouraged.
  – Self transitions are discouraged.

  Given two states i and j, and corresponding nodes on the topological map n_i, n_j, the transition $P([S_t = i] \mid [S_t - 1 = j]) = a_{ij}$ is expressed as:

$$a_{ij} = \begin{cases} a_1 & \text{if } \mathbf{U_T}[n_j] - \mathbf{U_T}[n_i] > 0 \ \wedge |\mathbf{U_T}[n_j] - \mathbf{U_T}[n_i]| > \epsilon \\ a_2 & \text{if } |\mathbf{U_T}[n_j] - \mathbf{U_T}[n_i]| \leq \epsilon \\ a_3 & \text{if } \mathbf{U_T}[n_j] - \mathbf{U_T}[n_i] < 0 \ \wedge |\mathbf{U_T}[n_j] - \mathbf{U_T}[n_i]| > \epsilon \\ a_4 & \text{if } \mathbf{U_T}[n_j] - \mathbf{U_T}[n_i] = 0 \end{cases} \tag{6.11}$$

[5] A high affordance implies easier access and thus lower cost while a low affordance implies more difficult access and a higher cost.

where $a_1, a_2, a_3, a_4 \in (0, 1]$ and ϵ is a small value. Due to the Delaunay Triangulation, in the Prior Topological Map, each node has at most three emanating edges. As a result, since the probabilities of all emanating edges from a state should sum to 1, Eqn. 6.11 needs to be normalised in the next step.

- ***Observation Probability:*** We follow the observational probability of a state as in (Pérez-Hurtado et al., 2015) and explained in Eqn. 6.3. Each state is associated with a separate Observation Probability represented as a Gaussian around it, centered on the position of its node. There is also a goal component associated with each state that contributes towards the observation probability. We follow here the representation proposed in Pérez-Hurtado et al. (2015) where it is assumed that pedestrians are moving in the direction of their intended goal, thus their velocity vector must point in the direction of the goal associated with the state. This probability is captured as a bivariate Gaussian $\mathbf{G}(\phi; \mu_\phi, \Sigma_\phi)$. The values of μ_ϕ and Σ_ϕ can be estimated based on the distance from the observation to the goal associated with the state.

6.3.3 Learning the Parameters

Learning the parameters, contrary to learning the topology and the structure of the graphical model which occurs at every observation, only happens at the end of the observation of the partial trajectory. Since the underlying graphical model of the GHMM is an HMM, an EM algorithm, in this case the Incremental Baum-Welch Method (IBW) method is used. The sensor model for the learning process is derived from Pérez-Hurtado et al. (2015).

6.4 Inference using Prior Knowledge

The aim of using Prior Knowledge in the form of a Prior Topological Map and associating the disparate links with transitions and priors derived from an observed environment, is to be able to infer pedestrian behaviour without the need for demonstrative trajectories. The generative model for the topology and, consequently, the underlying HMM of the GHMM should be able to describe an exemplar motion trajectory. With this in mind, given the inputs to this model are independent observations, taken with a camera or a LiDAR or any other observer, we expect an output to be probabilistic values over the states which can then be mapped into the subspace of the observation space $\Omega \mapsto \mathbb{R}^2$.

We predict motion the same way as described in the subsection 4.3.3. In essence, we perform prediction in three ways for an input sequence $O_{1:t}$ describing a partial trajectory. They are:

1. **State Estimation** A belief is maintained around every state of the GHMM. On observing a partial trajectory, this belief of the state is used to estimate the position of the pedestrian as in Eqn. 6.12

$$P(S_t \mid O_{1:t}) = \frac{1}{\eta} P(O_t|S_t) \sum_{S_{t-1}} P(S_t \mid S_{t-1}) P(S_{t-1} \mid O_{1:t-1}) \tag{6.12}$$

where η is a normalising constant, with the observation probability used being described in subsection 6.3.2.

2. **Goal Prediction** Over the course of a partial trajectory, it could be important to estimate the goal of the pedestrian under observation. Given the assumption made earlier (subsection 6.2.1) all the goals in the environment are known at observation time and thus, a belief over the goals for each timestep of the partial trajectory can be maintained. Since the state is composed of the node position and an associated goal, Eqn. 6.12 can be marginalised over all the nodes at each timestep to arrive at a probabilistic prediction as expressed in Eqn. 6.13.

$$P(p_t \mid O_{1:t}) = \frac{1}{\eta} \sum_{n} P(S_t = (n, \phi) \mid O_{1:t}) \tag{6.13}$$

Similarly, the node can also be found by marginalising over the available goals as in Eqn. 6.14

$$P(n_t \mid O_{1:t}) = \frac{1}{\eta} \sum_{\phi} P(S_t = (n, \phi) \mid O_{1:t}) \tag{6.14}$$

with η remaining the normalising value for both expressions.

3. **State Prediction** At every time instant of the partial trajectory, the purpose of state prediction is to find the reachable state at which the pedestrian can be found in, H timesteps in the future. This can be considered a propagation of Eqn. 6.12 over the states for a given horizon and can be expressed as in Eqn. 6.15.

$$P(S_{t+H}|O_{1:t}) = \sum_{S_{t+H-1}} P(S_{t+H} \mid S_{t+H-1}) P(S_{t+H-1} \mid O_{1:t}) \tag{6.15}$$

Consequently, predicting the most probable node after H timesteps can be performed by following Eqn. 6.14, for the state S_{t+H}.

To note, Eqn. 6.15 results in a probability distribution over discrete states. Instead, to compute the most probable observation in the observation space Ω at $t+H$ timesteps, the state prediction can be used as:

$$P(O_{t+H}|O_{1:t}) = \sum_{S_{t+H}} P(S_{t+H} \mid O_{1:t}) P(O_{t+H} \mid S_{t+H}) \tag{6.16}$$

6.4.1 Inference in Dynamic Environments

In this subsection, we present a further contribution of this thesis, of extending the GHMM method for prediction in dynamic environments. The motion of the pedestrian under consideration is modified based on the motion of other moving objects in the scene. This implies that at specific time instances some states in the graphical model are rendered unreachable due to the presence of obstacles at that state's position in the real environment. As a consequence, a concept of occupancy of the states can be introduced. This translates into the introduction of another variable Occ in the JPD as follows:

$$P(S_t S_{t-1} O_t Occ) = P(S_{t-1} Occ) P(S_t \mid S_{t-1} Occ) P(O_t \mid S_t Occ) \tag{6.17}$$

The occupancy variable, $Occ \in [0, 1]$, defines the behaviour of the graphical model, where $Occ = 1$ implies an occupied state and vice versa.

$$P(S_{t-1} Occ) = \begin{cases} 0 & \text{if } Occ = 1 \\ P(S_{t-1}) & \text{otherwise} \end{cases} \tag{6.18}$$

Similarly for transitions,

$$P(S_t S_{t-1} Occ) = \begin{cases} 0 & \text{if } Occ = 1 \\ P(S_t \mid S_{t-1}) & \text{otherwise} \end{cases} \tag{6.19}$$

Finally for the observation probability,

$$P(O_t S_t Occ) = \begin{cases} 0 & \text{if } Occ = 1 \\ P(O_t \mid S_t) & \text{otherwise} \end{cases} \tag{6.20}$$

From equations 6.18, 6.19 and 6.20, it can be inferred that when the state is occupied, it is equivalent to removing the state itself from the graphical model for that specific timestep. Since in a GHMM, these states have no free existence and are associated with corresponding nodes, the removal of states therefore implies the removal of these nodes on the topological map. In other words, inference is performed on a graphical model whose topology is limited by the existence of obstacles. This is exactly the function of the S-ITM whereby a topological map is generated disregarding those areas occupied by obstacles in the real environment. In essence, inference in a dynamic environment can be performed by the regeneration of the PTM at every timestep over the lifetime of the trajectory being predicted.

6.5 Discussion

In this chapter, we presented our extensions to the GHMM method to utilise Prior Knowledge for learning and inferring pedestrian motion. The first of these extensions was the Semantically-Aware ITM (S-ITM) algorithm which is aware of the properties of the semantics present in the environment while generating and updating the topology. This Prior Knowledge, once more, was used in the generation of the underlying graphical model and to provide accurate priors for this model. Finally, we presented the last of our contributions - the utilisation of the GHMM model for applications in dynamic environments.

The proposed method can be thought of as a mixture of the models proposed by Vasquez et al. (2009) and Pérez-Hurtado et al. (2015). Like the original implementation of Vasquez et al. (2009), each node in the topological map is associated with a uniform bivariate Gaussian distribution that provides for the observation model of the HMM. Also like the model of Pérez-Hurtado et al. (2015), in the proposed model, the Gaussian distributions are deformable under observations so as to be able to model pedestrian motion at a node more accurately via the observation model. Significantly, it differs from both in the sense that the proposed model, with its Prior Topological Map (PTM) acts in the same manner as a fully developed model of Vasquez et al. (2009) in structure while utilising partial trajectories to learn the parameters of the model with a sensor model equal to the one proposed in Pérez-Hurtado et al. (2015). This implies that the proposed method's PTM can be used to initialise a graphical model that can be used to infer pedestrian motion with no training trajectories.

Some limitations of this method are foreseen. For one, this method relies heavily on the accuracy of the generated cost map. A second limitation is the loss of learning during inference in dynamic environments. In this case of the proposed method, inference is performed at each timestep with the regeneration the PTM with no allowance for updating the topology. This could lead to the loss of valuable data or the missing of crucial trajectories in the topology.

In the coming chapters, we present a unified framework using the contributions of this and the previous chapter and testing the learning and prediction of pedestrian motion via the proposed method under different conditions.

Part III

Implementation and Experiments

7 Experimental Results - Exoperception

In this chapter, we present the results of experiments on our framework from the perspective of a static camera placed in an area overlooking the observation space containing a busy pedestrian crossing. Testing the framework in this configuration is important since it allows for complete observation of the environment. This implies that the motion of objects do not completely occlude one another, thereby presenting an opportunity to capture full trajectories throughout the environment of pedestrians, cars, and other shares space users and their interactions.

Here, we describe two classes of results - one in which we assume that there are no dynamic obstacles that affect the motion of the target pedestrian and the second, where dynamic obstacles are included in the inference and prediction process. For each of these cases, we provide qualitative and quantitative results under different cases of study with the purpose of establishing the accuracy and efficacy of the proposed framework under different conditions.

We commence by detailing the proposed framework and the constituent modules within. Next, we describe the dataset used for arriving at the results presented in this chapter. Using this chapter, we conduct the two classes of experiments *i.e.,* without and then with, moving objects. We finish this chapter with a discussion on the results.

7.1 Prior Knowledge Framework

In Chapter 5 we discussed a method to generate a potential cost map based on the different semantics observed in a scene within an urban area. Equally, in Chapter 6, we introduced an extension to the Growing Hidden Markov Model (GHMM) method designed to use a Prior Knowledge based cost map to predict pedestrian behaviour in an urban area. In this chapter, we construct a framework comprising the two separate approaches presented in the previous chapters to solve the problem of increasing the situational awareness of an autonomous vehicle in an urban area.

Consider an urban area where the observing agent $\mathcal{M}$ observes the real space Ω. Consider also the different semantics associated with regions within Ω such that $\Psi \subseteq \Psi^U$ and $\Psi^U = \{road, edge, crosswalk, sidewalk, O_{Static}, O_{Dynamic}, building, POI\}$.

Given such a configuration of real space, pedestrians are observed moving within it by the observer $\mathcal{M}$, possessing its own characteristics. This observer can either be a sensor mounted on an autonomous vehicle or placed on a building. We assume certain characteristics for this observer $\mathcal{M}$ such that:

- it is completely known - i.e., the different parameters of the sensors are known in advance.
- it is discreet - it does not affect the behaviour of the subjects being observed[1].
- it can localise and identify the semantics of the region around itself.
- it can identify and track pedestrians and other dynamic obstacles.
- it can estimate the geometric dimensions of the regions associated with the different semantic labels - for example, knowing the width of an urban road.

The knowledge by the observer as listed above is considered Prior Knowledge on which the rest of the framework is based.

On the acquiring this Prior Knowledge, the framework should be able to answer the following questions:

- **(I1)** Knowing the semantic morphology of the observed space Ω, is it possible to predict future positions of observed pedestrians along with their possible goals?
- **(I2)** On knowing the semantic morphology and previous observations, can typical and atypical trajectories be explained and thus, future states of the observed pedestrians predicted?

(I1) represents the problem of representation and prediction in an novel, previously unobserved area while **(I2)** represents the traditional learn-and-predict problem for pedestrian trajectories.

In Section 1.3, a general architecture of a framework that models the environment, learns behaviour and predicts pedestrian motion was described. Here, we present an expanded version of the same architecture within the specific context of predicting pedestrian motion in built, urban areas.

Fig. 7.1 depicts the proposed framework and the interaction between the various components within. The global input to the framework is predominantly the sensor data and data received from a tracker, tracking all the objects within the observed scene. The red blocks in the diagram represent the main contributions of this thesis. In the coming subsections, we will describe each of these modules in detail.

[1] This becomes especially important on an autonomous vehicle since studies (Rothenbücher et al., 2016; Ferrier-Barbut et al., 2018) have shown that the behaviour of pedestrians around self-driving cars is heavily impacted.

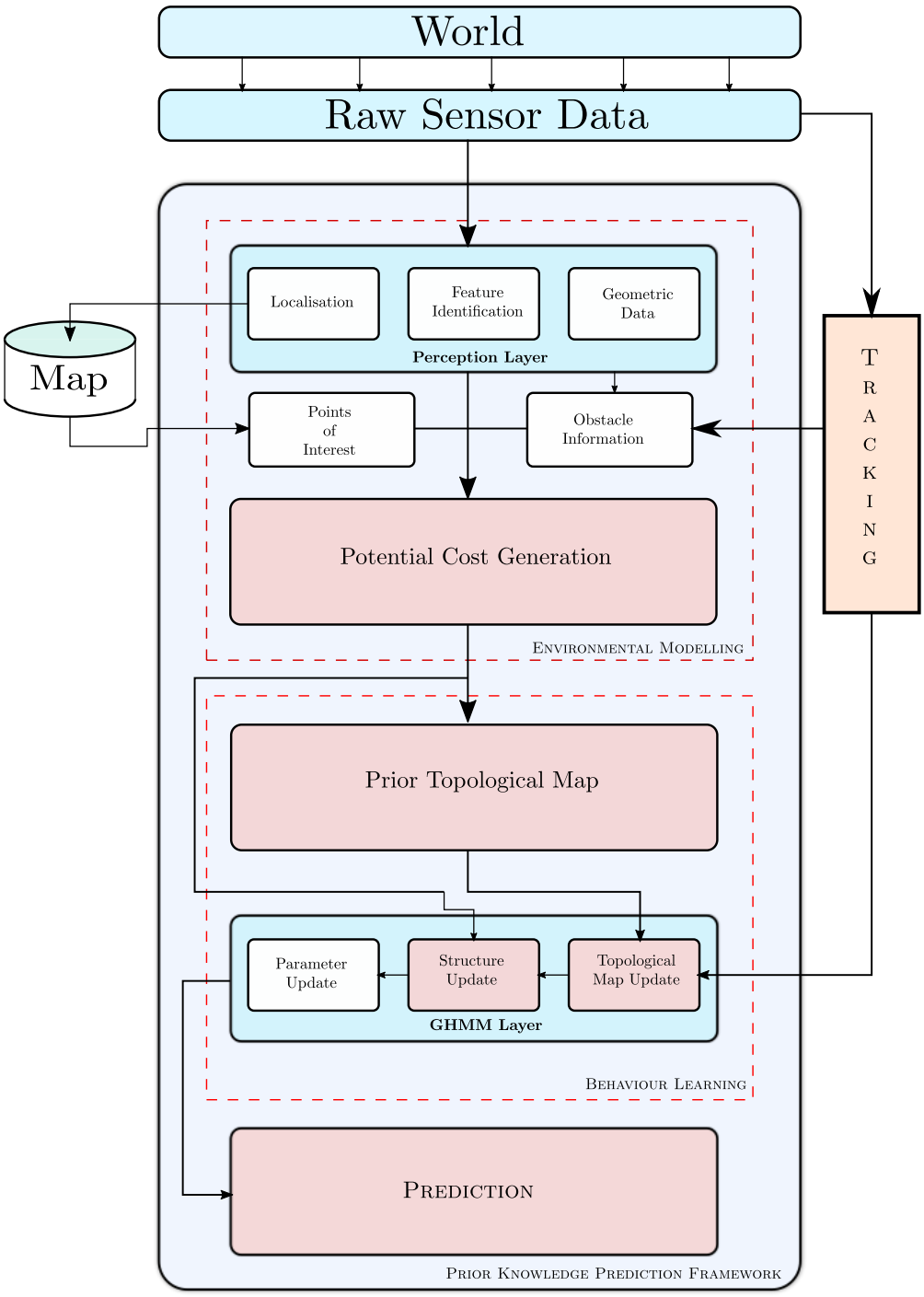


Fig. 7.1.: Architecture of the proposed framework with its constituent modules marked out in red, dashed lines. The blocks in red depict the innovations presented in this thesis.

7.1.1 Environmental Modelling

Building the cost map is the responsibility of the Environmental Modelling module in the proposed architecture. First, various proprioceptive sensors pick up data about the world. This data is then parsed through the "Perception Layer" seen in Fig. 7.1. Here, in this layer, the raw sensory input is converted into information that can be utilised for the successive modules of the framework. Within the "Perception Layer", incoming data is parsed through a system to identify the different features of the observed environment. Secondly, geometric information about the environment (*e.g.,* width of the road, edge thickness, *etc.*) is estimated for the identified features. Next, in localising the observer and passing this information through a map, relevant POIs are identified in the environment. Finally, information regarding the dynamic obstacles in the environment are received through an external component in the form of a tracker using the raw data input from the world.

This information, once acquired, can then be parsed through the "Potential Cost Generation" module to generate a cost map of the environment via algorithm 5 presented in Chapter 5.

Thus, as a summary, the Environmental Modelling module performs the following functions -

- Identification and localisation of the different POIs in the environment and their relative importances.
- Classification of the observed scene into the different semantic features Ψ,
- Use of the geometric data of the scene to get a measure of the width of the street, length of the sidewalk, etc.
- Tracking pedestrians within this scene.
- Generating the Potential Cost Map.

From a practical perspective, the different Ψ changes depending on the context. For example, the Ψ for an urban street is different from that of a parking lot as is the relative importance of the POIs in the scene (refer Eqn. 5.14).

In the following section describing the use of different datasets to validate our framework, we detail the methods used to estimate and identify the contexts itemised above. One remark to be noted here is the absence of a direct connection between observed trajectories and the generated cost map. It can be observed that observed trajectories of dynamic obstacles (pedestrians, cars) are only utilised for information about their motion and not directly as input to the module to learn weights as in some of the previously presented literature.

Another note to be made at this junction is the importance of the "Perception Layer" in the presented module. Good identification and tracking algorithms play a crucial role

generating accurate cost maps in our framework, the absence of which will lead to bad predictions from the associated modules.

7.1.2 Behaviour Learning

The second module in the pipeline of the proposed framework is the Behaviour Learning module. This module performs two tasks -

1. Capture pedestrian trajectories in novel and previously unseen areas
2. Learn new trajectory prototypes based on witnessed trajectories.

We perform these tasks by utilising the generated cost map from the Environment Modelling module and observational data from the tracker. Primarily, we generate the "Prior Topological Map" in the module of the same name by following the procedure set out in subsection 6.2.1 on the potential cost map from the previous stage. This topology is then fed to the "GHMM layer" to instantiate the graphical model with appropriate values derived from the cost map and the nodes and edges of the prior topological map. Consecutively, this graphical model becomes sufficient to perform task 1.

Task 2 necessitates the inclusion of pedestrian observations provided by the tracker in the form of *Augmented Observations*. These observations modify the Prior Topological map represented by $\mathcal{E}$ in accordance with the S-ITM algorithm. Consecutive observations in the form of partial trajectories from the tracker further update the topology, structure and the parameters of the underlying HMM following the material presented in Section 6.3. In learning the topology and parameters derived from new trajectories, the graphical model is adaptable to capturing non-typical trajectories better.

The output of this module is time dependent which is a tuple $\mathrm{X}_t = \langle \mathcal{E}, \lambda \rangle$ of the topological map $\mathcal{E}$ and the Hidden Markov Model λ, each with their constituent parameters that have been learnt using the partial trajectories.

7.1.3 Prediction

The Prediction module, as evident, deals with task of pedestrian prediction utilising the GHMM model generated from the previous component. This prediction is a naive Bayesian inference for a time horizon H with the beliefs maintained over the GHMM graphical model. Prediction, in our chosen context, can be performed under two scenarios - by including or excluding the dynamic objects in the observed scene. In the former, these dynamic objects play an integral part in the prediction process.

In the first scenario, a graphical model from the previous component, built with only static objects, performs inference at time t on receiving a trajectory (partial or full) from

the tracker utilising the tuple X_t for predicting pedestrian behaviour. Since the framework is a lifetime learning-predict system, prediction is performed on the latest iteration of X.

The second scenario is dependent on information about the different dynamic objects in the scene. This information first needs to be captured by the perception layer and then tracked by the tracker. Information on, for example, the velocities of these objects need to be passed to the potential cost map module to generate the corresponding cost map at that instant of the environment. A prior topological map on this cost map is generated next and it is this topology that instantiates the underlying HMM in the GHMM layer within the architecture. On this iteration of X, inference for the trajectory being tracked is performed. On the arrival of a new cost map due to the changing nature of the environment, a new time-dependent X is generated and utilised for prediction with the old one being discarded.

7.2 Dataset Description

The results presented later in this chapter have been conducted on the dataset presented below. A large problem facing current research, one that we experienced during the course of this work, is the dearth of comprehensive pedestrian datasets in urban areas (Ridel et al., 2018). A main criticism in using a dataset such as the Daimler Pedestrian Dataset (Schneider and Dariu M. Gavrila, 2013) is that the pedestrian behaviour is not natural i.e., actors perform certain predefined pedestrian tasks. For this reason, we annotated existing videos of pedestrians (subsection 7.2.1) and collected data on pedestrian behaviour using the experimental platform in specific environments such as a parking lot.

In the previous section, we discussed the importance of the "Perception Layer" within the framework. Thus, for each of the datasets detailed below, we present the methods used to:

- Identify the configuration of the observed environment
- Estimate the geometry of the environment
- Identify the obstacles – static and dynamic
- Extract pedestrian and other shared space user trajectories.

7.2.1 Traffic Anomaly Dataset

The Traffic Anomaly Dataset is a video which we annotated with the original video taken from Varadarajan and Odobez (2009). This video is captured from a camera featuring (Fig. 7.2) a busy urban street containing a crosswalk, a sidewalk, the street, a Point of Interest in the form of a restaurant and an unseen Point of Interest directly below the camera in the form of a commercial center. Specifically, from a semantic viewpoint, the scene

has a clear demarcation between the sidewalk and the road with visible edges to the street. Secondly, it has a traffic light regulating the pedestrian crossing as well as driving cars and other shared space users, making it a well structured environment.

This configuration of the environment is extremely interesting since this is a typical shared space urban environment. In the video pedestrians are observed to behave naturally in such an environment, with many waiting at the red light to cross while others decide to cross illegally. Many more decide to traverse the length of the scene. Others enter the restaurant present in the scene.

Four points of interest can be identified in the scene using the GPS coordinates of the observing camera. Three of them are the edges of the scene where pedestrians appear and disappear from and the last being the restaurant in the view. These POIs are depicted in Fig. 7.2 as red, numbered dots.

Fig. 7.2.: Still from the Traffic Anomaly Video (Varadarajan and Odobez, 2009). The environment contains a crosswalk, a visible sidewalk, edges and road. It also contains four Points of Interest that have been numbered in the image.

From this environment, we have extracted real data trajectories as explained in the rest of this section.

Extraction of trajectories

To extract trajectories from this video, we first attempted automatic tracking of dynamic objects in the scene using the YOLOv2 (Redmon and Farhadi, 2016) and DeepSORT (Wojke et al., 2017) pipeline on the video. The YOLOv2 architecture processes raw frames as input and produces bounding boxes for “pedestrian”, “cycle”, “car” and “truck” classes. These identified objects are then passed to the DeepSORT architecture to track them over time. The issues we faced here were twofold - a) there was a proliferation of dynamic obstacles within the video and b) the resolution of the video was small - leading to the problem of

Data Association. The tracker was unable to reliably maintain a tracking ID on many pedestrians over the course of their trajectories, especially when there were multiple objects in the scene. To solve this association issue, we used the VATIC tool (Vondrick et al., 2012) to manually annotate and correct trajectories so as to arrive at reliable full trajectories for every "pedestrian" class in the video.

Post processing involved the extraction of the position of the feet from the tracked pedestrians. Velocities over the course of the trajectory were obtained by running them through a Kalman filter.

All pedestrian trajectories extracted from the video are presented in Fig. 7.3. In comparison to the physical topology presented in Fig. 7.2, the behaviour of pedestrians in this scene can be clearly witnessed. Fig. 7.3 clearly shows the preference of pedestrians to complete their trajectory in safe (*also:* typical, "legal") paths. Conversely, there are other non-typical ("illegal") trajectories present in the dataset wherein the pedestrian wishes to cross to the other side over the unmarked part of the road, which is inherently unsafe.

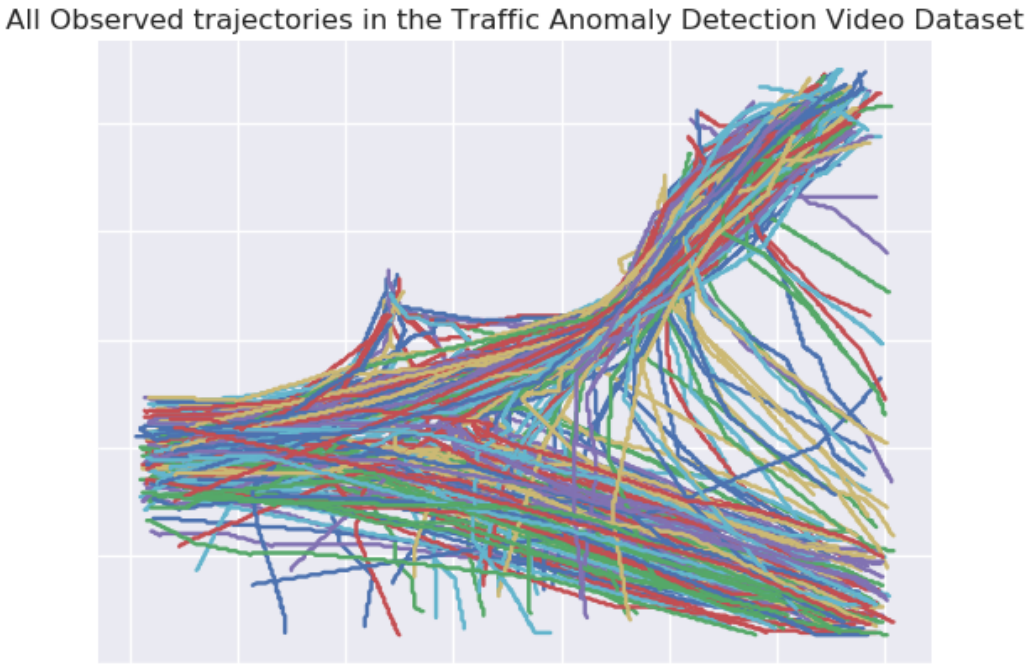


Fig. 7.3.: Tracks of all pedestrian trajectories captured from an overhead, stationary camera at in an urban area for the Traffic Anomaly Dataset. The tracked pedestrians exhibit typical and non-typical behaviour at this traffic junction.

Semantics and Geometry

While the dynamic objects in the scene are easily identified using YOLOv2, the other elements of the scene needed labelling. On one hand, automatic labelling of these features can also be performed using methods proposed in (Munoz et al., 2010), or use neural network

based methods like (Badrinarayanan et al., 2015). On the other hand, these networks rarely perform well when the image to be inferred are not similar to the data they are trained on. Thus, off-the-shelf solutions may not be reliable. To avoid this problem of mislabelling, we manually labelled the different features of the scene (i.e., the road, sidewalk, crosswalk, edge and building). To find the geometric parameters of the scene, we estimated the camera parameters and utilising this, we found the homography of the image. The width of the road was estimated using the GPS coordinates of the observing camera and the Swiss road dimension standards.

7.3 Experimental Results - Exclusion of Moving Objects

This section deals with the first of the classes of experiments, wherein the moving obstacles within the scene are disregarded in favour of a static scene to gauge the accuracy of the proposed framework in comparison to the state of the art method found in literature in the field of GHMMs. This state of the art method, proposed in Pérez-Hurtado et al. (2015)[2], has already been treated in Chapter 6. Since that method disregards dynamic objects in the inference process, the experiments conducted here also follow the same principle, for fair comparison. This comparison is performed along two separate dimensions - the first dimension of comparison is across different epochs of training for the competing models. The second dimension of comparison is across pedestrian behaviour in a structured environment, namely that of legal and illegal crossings. Thus, all the results presented within this chapter will be compared against these two dimensions with analyses on the obtained results.

The method used for the experiments follows the pipeline presented in the Prior Knowledge Framework (Section 7.1) and is as below:

1. The potential cost map for the static scene will be generated as proposed in Chapter 5 and 5 for the scene depicted in Fig. 7.2
2. For this scene, utilising the potential cost map, the prior topological map is generated.
3. This topological map is then converted into the underlying graphical model via the GHMM method, resulting in the nodes and edges of the topological map generating their corresponding states and transitions.
4. Partial trajectories, derived from the training set of the dataset presented in subsection 7.2.1 are used to update the topological map so as to learn atypical trajectories and pedestrian behaviours in this environment. These updates to the topological map also correspondingly update the underlying graphical model's structure.

[2]From hereon in called "competing method"

5. The partial trajectories that update the topology are used to learn the parameters of the GHMM.

6. At different epochs, testing trajectories - divided into “legal” and “illegal” trajectory sets are used to predict pedestrian positions at a specified horizon.

7. These predictions are then compared against the ground truth of the testing trajectories availed from the dataset.

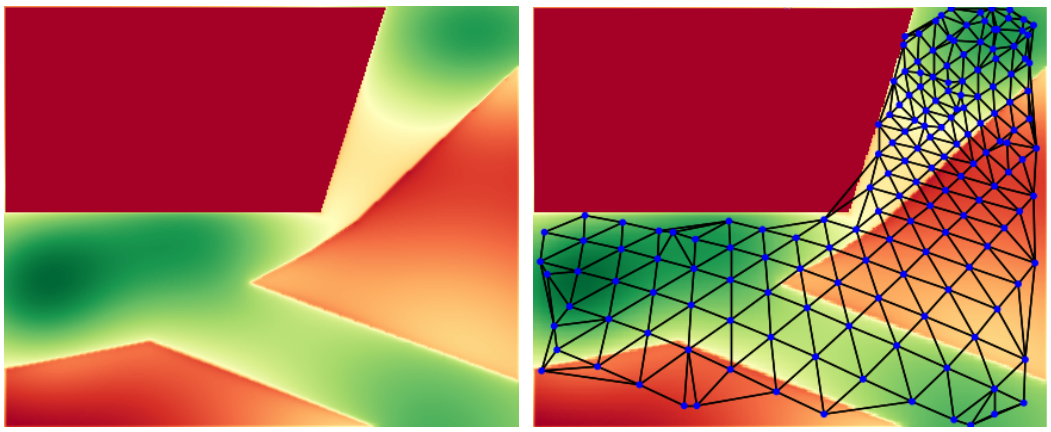
On the contrary, the model compared against does not privilege the use of the prior topological map and its corresponding changes. Thus, instead, for comparison, the training of this model including the topological map updates are performed sequentially as presented in steps 4 – 7. Significantly, the input of the partial trajectories to both models happen simultaneously such that they are both trained on the same observations. We then present the qualitative and quantitative results for this class of experiments.

Implementation details

For the dataset in this specific scene, different parameters were obtained via various means. These can be separately classified as the Potential Cost Map parameters and the GHMM parameters. The first set of parameters that were necessary but not directly utilised in the implementation of the framework, were the camera parameters. Since we only annotated the original video and not capture it, we were unable to find the camera parameters of the capturing device. To do this, we chose different points at regular patterns (such as the edges of crosswalk hatches whose lengths and widths are standardised) and assumed average heights for pedestrians. Using these points, we were able to approximately estimate the intrinsic and extrinsic parameters of the camera. These values were used to generate a topological grid that conformed to the contours of the observed environment instead of a flat, 2D structure of the image.

The width of the road, L_{Road} was found to be $9m$, the edge parameter η is arbitrarily set to a value higher than the road parameter. All the destinations in the scene are assumed to have equal attractiveness for pedestrians. The insertion distance τ for the topological map is set to a value of 1.5 meters. On the other hand, for training the competing model, while τ remained the same, the default prior value π_0 and the default transition value a_0 on the creation of states and transitions are arbitrarily chosen to be 0.5 and 0.5 respectively.

Using these parameters, the potential cost map and the prior topological map are generated and can be observed in Fig. 7.4. These values and the graphical model obtained via these two depicted images will be used to obtain the rest of the results presented over the rest of this chapter.



(a) Figure showing the generated potential cost map based on a labelled image of the dataset. This potential cost map only depicts the static environment and does not take into account any motion. As a gradient from green to red, greener areas are those with lower affordances while red regions signify high affordances. Dark red regions in the scene correspond to obstacles, in this case the building seen in Fig. 7.2.

(b) The Prior Topological Map generated based on the potential cost map depicted adjacently. The nodes and the edges depict reachable areas in the observed environment.

Fig. 7.4.: Potential Cost Map and Prior Topological Map for the observed environment in the Traffic Anomaly annotations dataset.

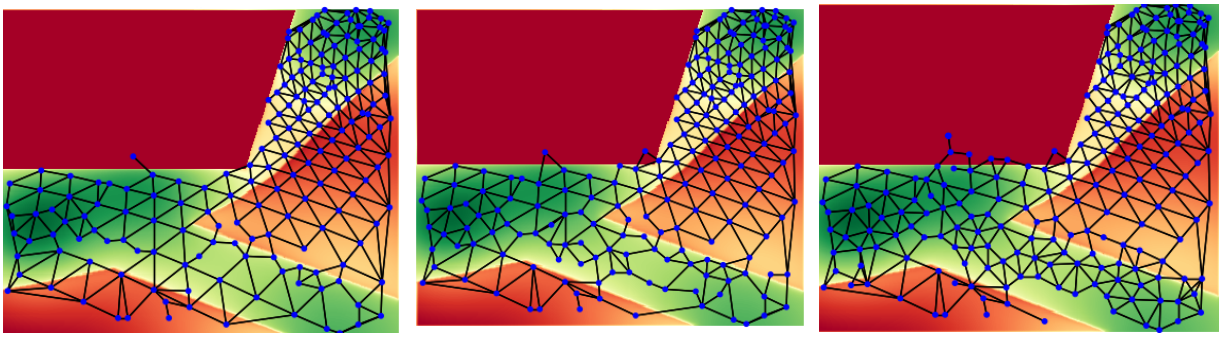
7.3.1 Qualitative Results

The ultimate purpose of the proposed Prior Knowledge Framework is for the observing agent, be it a static observer with a bird's eye view of the environment or a moving platform, to be able to predict pedestrian trajectories under different conditions of observation or behaviours in novel environments. Consequently, the quality of this prediction becomes tantamount to the model's ability to accommodate all behaviours in different conditions.

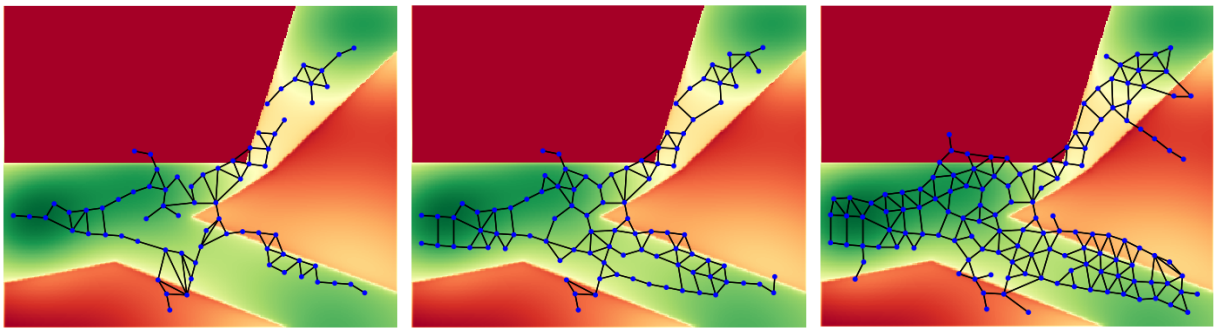
The criterion for this qualitative evaluation, in our context, is to "cover" more of the observed area. Considering that the nodes and edges of the topological map can be directly translated as paths that a pedestrian can take to move from one region of the environment to another[3], and that for the model to predict an atypical trajectory with each observation of this trajectory considered accommodated if it can be associated with an already existing node, it is logical to accept that having more areas of the observed space discretised by these nodes (and connected via edges) can accommodate these trajectories better. More formally, for every observation of the trajectory under inference, if there exists a node n within the insertion criteria defined by the insertion distance τ, then the observation can be considered "accommodated" within the existing topology.

[3] refer Chapter 4

Following the experimental procedure described earlier, we attain the qualitative results seen in Fig. 7.5. Fig. 7.5a demonstrates the evolution of the topological map over different training epochs via the proposed method with the epochs being 50, 100 and 250 partial training trajectories drawn from the dataset described in subsection 7.2.1. It is noticed that this topological evolution is initiated from the prior topological map. On the other hand, Fig. 7.5b provides the evolution of the topological map for the same training epochs with the same trajectories, for the competing method.



(a) Evolution of the topological map under the proposed method for 50, 100 and 200 training partial trajectories. This evolution takes into account the environmental context during the update phase.



(b) Evolution of the topological map under method proposed in Pérez-Hurtado et al. (2015) for 50, 100 and 200 training partial trajectories. The evolution of the map does not take into account the environment.

Fig. 7.5.: Qualitative comparison in the evolution of the topological map between the proposed approach and the competing approach.

Consider the three trajectories depicted in Fig. 7.6. Each trajectory exhibits a characteristic behaviour witnessed in an urban environment. The blue trajectory depicts a fully legal trajectory by transiting from one POI to another within areas designated as legal. Likewise, the red trajectory, over its lifetime, exhibits fully atypical behaviour by transiting across the road area of the observed environment. On the other hand, the green trajectory exhibits a mixed behaviour with a substantial part of its lifetime remaining in the legal areas of the environment, thus performing typical actions while its behaviour in some parts, near the end of its lifetime, coincides with that of an illegal action. These specimen trajectories can

be utilised to explain the inference process of the proposed framework, thereby presenting a qualitative and quantitative evaluation of this framework under the specified conditions.

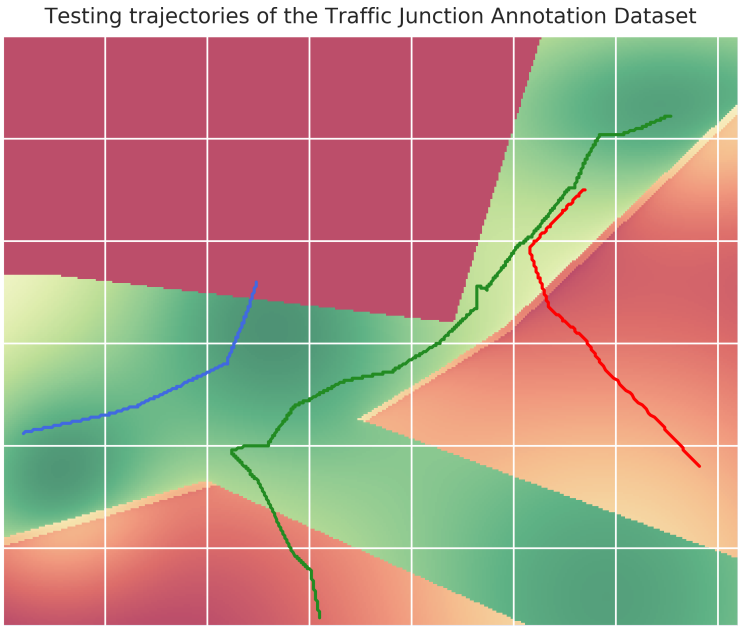


Fig. 7.6.: Specimen test trajectories chosen from the Traffic Anomaly Annotation Dataset. The three specimen trajectories exhibit three different behaviours in the environment. The blue trajectory is a fully legal trajectory. The red trajectory is a fully illegal trajectory and the green one exhibits mixed behaviours with a significant portion of the trajectory behaving typically while the remaining part exhibits atypical behaviour.

For the fully legal trajectory, with the proposed model at different training epochs, there always exists atleast one node within the insertion criteria. Thus, all observations emanating from this trajectory can be accommodated within the existing topologies. Similar conclusions can be deduced for the illegal and mixed behaviour trajectories depicted above. Conversely, with regard to the competing model, it can be observed that there exist no or few nodes across the different training epochs to accommodate the illegal trajectory. Meanwhile, for the mixed behaviour trajectory, the accommodation gets progressively better over training epochs just as for the legal trajectory.

7.3.2 Quantitative Results

For quantitative results, as an evaluation criterion, we measure the accuracy of prediction over the lifetime of the testing trajectory under the same training conditions as above. The error between the predicted position and the ground truth is used as a measure to discriminate between the performance of the proposed and the compared methods.

The procedure for obtaining these errors remains the same as earlier. The underlying models for the proposed method and the competing method are trained for the very same

training epochs as in subsection 7.3.1. The purpose of this experiment is to illustrate that the proposed method requires fewer training data than the competing method without sacrificing accuracy.

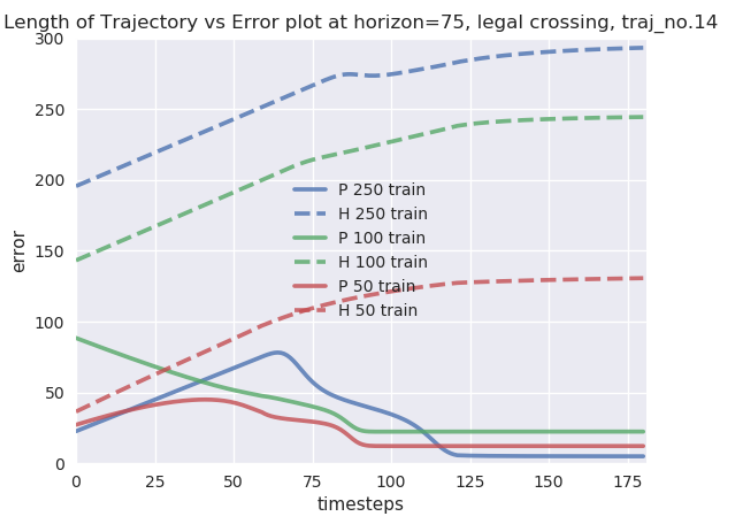
A second experiment is conducted characterised by prediction performed on an untrained model of the proposed method (*i.e*, the underlying graphical model initialised with the Prior Topological Map) and compared against the competing model that has been trained with all available partial trajectories. This experiment simulates a scenario where the autonomous vehicle enters an area wherein it cannot wait in the area to gather many partial trajectories before performing inference.

Under these conditions, the prediction errors are obtained and plotted as seen in Fig. 7.7. Figures 7.7a and 7.7b represent the prediction errors for the legal and illegal testing trajectories for different training epochs. Figures 7.7c and 7.7d represent the second experiment pertaining to a fully trained model *vs.* an untrained proposed model.

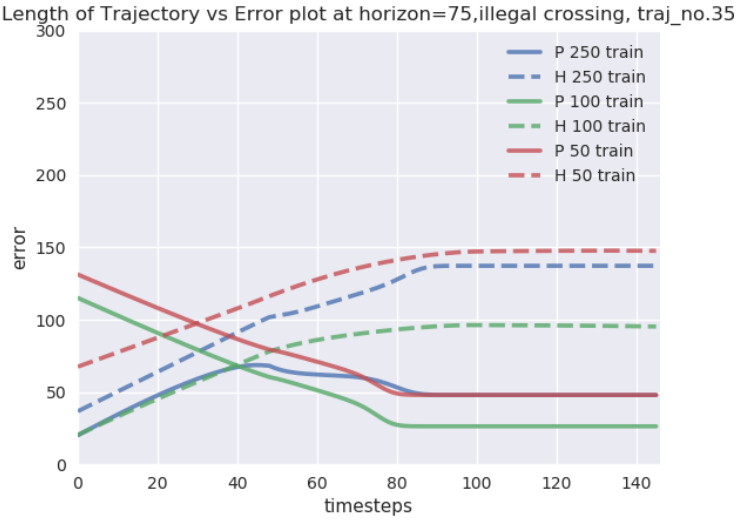
The results portrayed by figures 7.7a and 7.7b are in line with the discussion on the qualitative analysis of these trajectories. It is observed that for the proposed method, the errors quickly converge to very small values, proving that there are accommodating nodes along the entire trajectory with the implication being that there exist corresponding states and transitions that provide for accurate inferences over short and medium horizons. On the contrary, for the competing method, with the same logic, the prediction error remains high throughout the lifetime of the testing trajectories over the different training epochs.

The results obtained in figures 7.7c and 7.7d where it is observed that the proposed model with no training performs better than a fully trained competing model for both typical and atypical trajectories can be explained on the basis of availability of training partial trajectories. Returning to Fig. 7.3, we can observe that most trajectories are heavily concentrated around the legal areas. The test trajectories were so chosen as to be outliers, even amongst the legal class of the dataset. Secondly, the training partial trajectories fed to the models are not sequential. Thus, the model does not receive, say two partial trajectories comprising the same full trajectory in sequence. Instead, it may receive two completely unrelated trajectories in sequence. This might lead to overfitting of the data in some regions while starving other regions of the environment. Since the proposed model with no training is associated with a uniform Gaussian around each state of the model at initialisation, the above-mentioned problem does not arise thereby leading to better performance than the competing method.

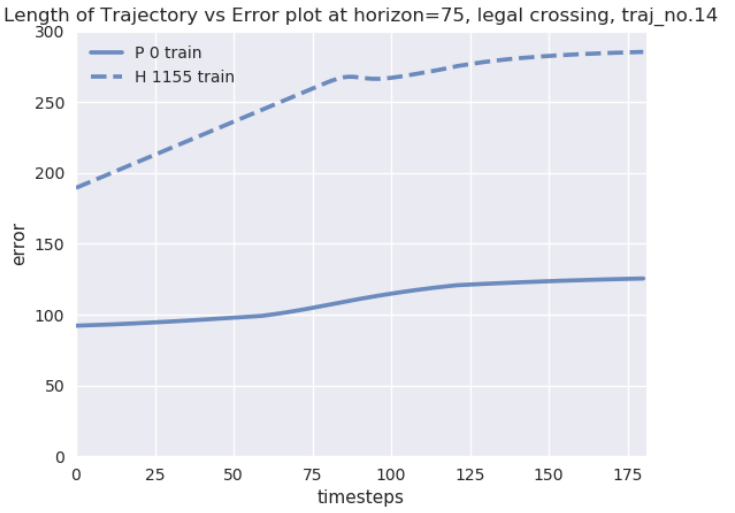
This is not to say that the proposed method always performs better. Let us consider the performance of the mixed behaviour trajectory under the second experiment. This result is depicted in Fig. 7.8. Here, the proposed method performs worse than the competing method in the former half of the trajectory's lifetime and in the latter, performs better. The prediction error of the competing method rises slightly in the latter part of the trajectory. With this result, we can comment on the utilisation of the two methods.



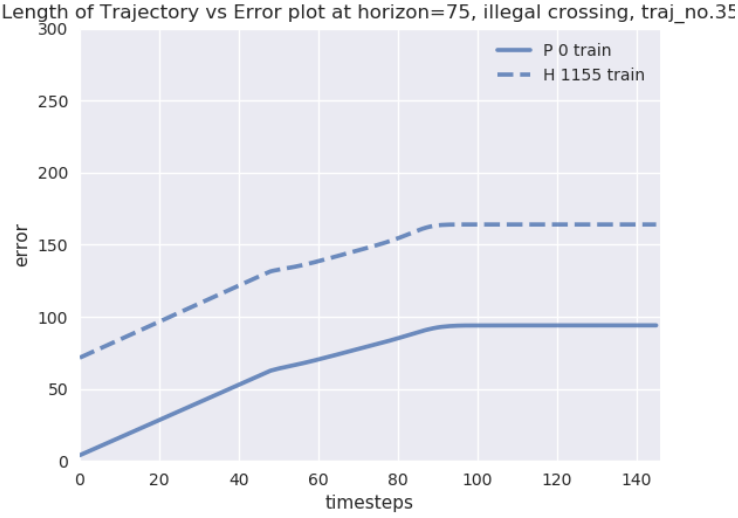
(a) Prediction accuracy for the legal trajectory at different training epochs over its lifetime.



(b) Prediction accuracy for the illegal trajectory at different training epochs over its lifetime.



(c) Prediction accuracy for the same legal trajectory as above tested under full training for the competing method and no training for the proposed method.



(d) Prediction accuracy for the same illegal trajectory as above tested under full training for the competing method and no training for the proposed method.

Fig. 7.7.: Quantitative results for the legal and illegal trajectories. The upper row depicts the prediction errors for the tested trajectories over different training epochs for the proposed and the competing method. Solid lines (annotated with P in the legend) pertain to the proposed method. Dashed lines (annotated with an H in the legend) pertain to the competing method. Colours for a training epoch remain the same over the two methods. The bottom row depicts prediction error for the same trajectories with a fully trained competing method plotted against the proposed method. As previously, solid lines refer to the proposed method while the dotted ones to the competing method. The error is in pixel distance.

1. The competing method performs better inference during the typical (legal) part of the trajectory. Specifically, this dataset and the method in which the training was performed explains this result. Data is overfitted along the sidewalk region of the environment, where most of the training data is concentrated leading to better prediction of positions at this part of the trajectory. The proposed method, at 0-training epoch, has no such bias which is both a boon, an in this case, a curse.

2. The proposed method performs better than the competing method during the atypical part of the trajectory. Conversely, the competing method can be generalised to say that for this dataset where typical behaviours far outnumber atypical ones, the competing model suffers to perform accurate inference.

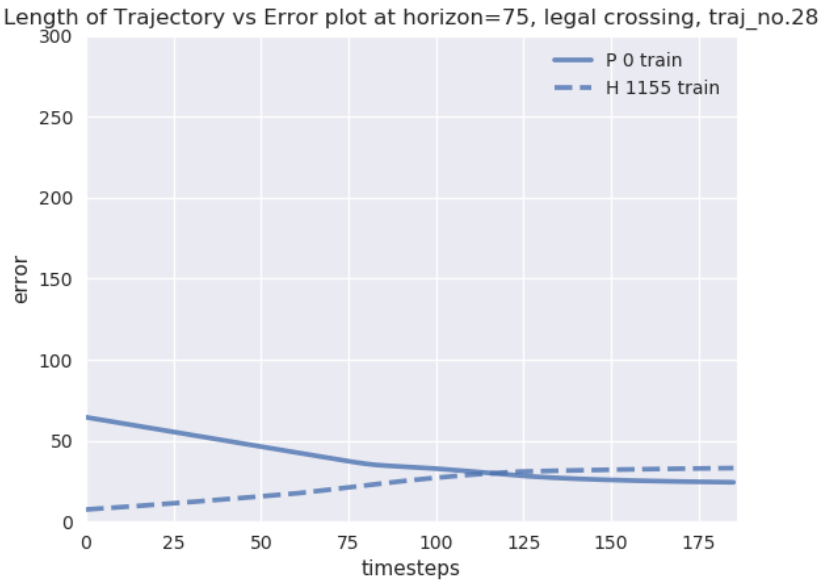


Fig. 7.8.: Quantitative comparison for the mixed behaviour trajectory depicted in Fig. 7.6 at no training of the proposed model (in a solid line) *vs.* full training of the competing model (in a dashed line). It is observed that for a part of the trajectory the competing method performs better than the proposed approach.

To check whether our model performs significantly better compared to the competing model, we performed a t-test on comparing prediction errors from all the trajectories in the "Legal" and "Illegal" subsets of the dataset. The p-values for this test is given in Table 7.1.

All values are above our significance threshold of 0.05. As a consequence, we can surmise that the proposed model *does* perform better with significant accuracy against the competing approach over a variety of trajectory classes.

Trajectory type	Number of learning (partial) trajectories			
	50	100	250	0 - Full
	p-value	p-value	p-value	p-value
Legal	7.023×10^{-16}	1.314×10^{-18}	2.373×10^{-39}	2.172×10^{-61}
Illegal	5.602×10^{-33}	3.006×10^{-33}	1.178×10^{-29}	4.907×10^{-21}

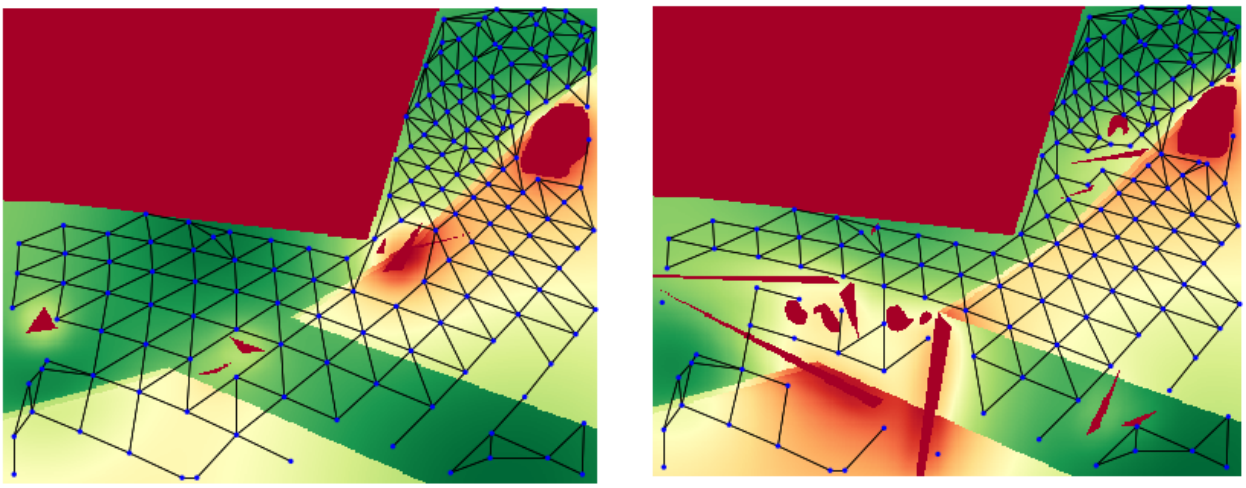
Tab. 7.1.: Comparison of prediction accuracy at horizon=75 (corresponding to $3s$) at varying training levels between the proposed work and Pérez-Hurtado et al. (2015). Lower values are better.

7.4 Experimental Results - Inclusion of Moving Objects

This section deals with the second class of experiments with the Prior Knowledge framework characterised by the inclusion of moving objects into the inference pipeline. In comparison with the experiments conducted excluding the moving objects in the scene, there are some similarities and differences in the procedure. The differences stem in view of the fact that there is no training phase of the GHMM. Instead, the framework acts only under the influence of the generated Prior Topological Map and consequently the underlying graphical model for this topology. Ideally, given a continuously changing environment, there should a dynamic potential cost map with topological maps being drawn on them at every timestep over the lifetime of the trajectory under observation and inference. Practically, however, we perform these experiments by freezing the state of the environment at a specific timestep and feed the feature details derived from this environment state into the framework. From that point on, the rest of the procedure remains the same for testing the prediction accuracy as presented in the previous section.

To present the results of this experimental class, we utilise the fully legal and fully illegal trajectories presented in Fig. 7.6. First, we generate the potential cost map for both the specimen trajectories. This is accomplished by freezing the state of the environment at timestep t_1, the first timestep after the commencing of the trajectories. The decision to freeze the environment is deliberate. This allows us to identify the objects present at the beginning of the trajectory *i.e.,* at t_0 and consequently, estimate their velocities at t_1 as required for their modelling. The other features of the scene are identified, along with the POIs and the potential cost map is generated. On this potential cost map, the Prior Topological Map is drawn using the S-ITM. These topologies are as depicted in Fig. 7.9.

Qualitatively, we use the same criterion used for the experimental class excluding moving objects. For a legal trajectory starting from the bottom left of the environment and reaching the POI such as the specimen trajectory, the moving objects with the costs inherent to their motion have no adverse effect on the tracked trajectory based on the topology generated and depicted in Fig. 7.9a. On the contrary, consider the state of the environment and the topological map depicted in Fig. 7.9b. This figure presents an interesting scenario for a



(a) State of the world at t_1 for the legal trajectory under consideration.

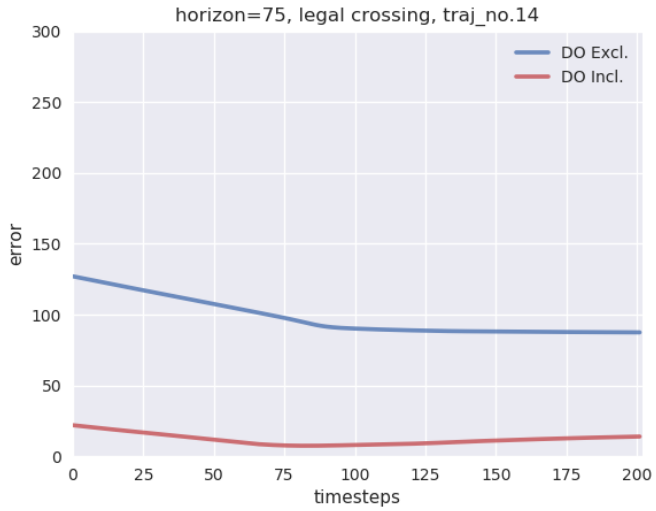
(b) State of the world at t_1 for the illegal trajectory under consideration.

Fig. 7.9.: Prior Topological Maps drawn for the state of the environment at timestep t_1 for the specimen trajectories. The potential cost map introduces costs derived from the velocities of the moving objects in each of the scenes.

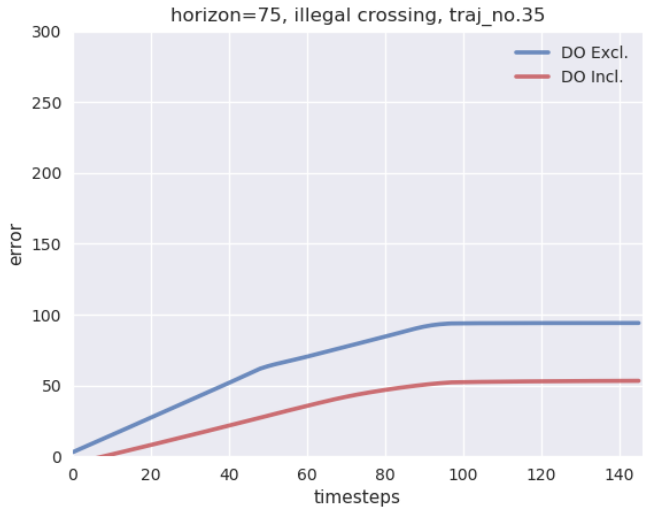
trajectory originating at the bottom right of the image. The topological map and the motion of other objects in the scene forces this trajectory to veer off the typical areas with the implication being that no legal crossing is possible from that origin. Indeed, performing a crossing demands motion in areas of higher potential cost.

The quantitative evaluation for the testing trajectories corresponding to the topological maps presented for the legal and illegal trajectories above can be seen in Fig. 7.10a and Fig. 7.10b respectively. We observe, for both specimen trajectories, that the introduction of moving objects increases the accuracy of prediction. This effect occurs because the moving objects reduce the number of reachable nodes for the target trajectory during the inference process and constrict the probabilities of these states, increasing the prediction accuracy.

While in the demonstrated examples the introduction of moving objects increases the accuracy of prediction, there are examples to the contrary. Such an example for this behaviour can be seen in Fig. 7.11. This figure depicts the prediction error evolution of a legal trajectory for including and excluding moving objects. As can be noticed, the inclusion of dynamic objects in the inference process leads to a sudden increase in the prediction error only to decrease towards the end of the trajectory. This behaviour of the model can be explained by the procedure followed in performing the prediction. Here, the state of the environment is frozen at t_1 of the trajectory and prediction performed under this constraint. Significantly, this assumption is unable to include into the inference process, the interaction between the pedestrian being tracked and the other moving object. This interaction could lead to sudden, short-term divergences from typical behaviours. Once this disruptive ob-



(a) Prediction accuracy for the legal specimen trajectory on untrained models with and without dynamic objects.



(b) Prediction accuracy for the illegal specimen trajectory on untrained models with and without dynamic objects.

Fig. 7.10.: Quantitative comparison of prediction accuracy of the proposed model excluding and including dynamic objects for the specimen trajectories. The accuracy over the lifetime of the trajectory when the framework excludes and includes dynamic obstacles have been depicted in solid blue and red lines respectively. Since the competing approach is not intended to be utilised in dynamic environments, we compare prediction accuracy against our own method excluding dynamic objects on untrained models.

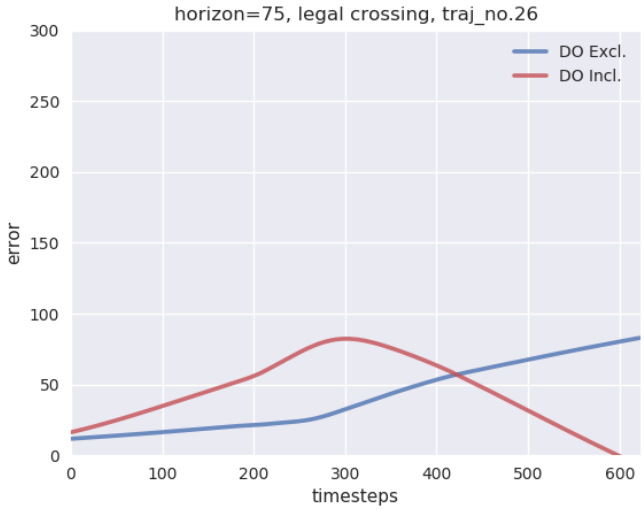


Fig. 7.11.: Prediction accuracy comparison for a legal test trajectory predicting in an environment with dynamic obstacles performing poorly against a model inferring in their absence.

ject has been passed, the pedestrian might choose to reconverge to his original behaviour or choose to take an entirely different path.

7.5 Discussion

In this chapter, we presented the results obtained on the proposed framework via experiments conducted on a real dataset under different conditions, from a static observer overlooking the observed area. These conditions were the inclusion and exclusion of moving objects in the inference process; the efficacy of the framework when compared to a competing model at different training epochs and significantly, proving that our framework performs better than the competing model without the need to train.

The latter two conditions specifically are in line with the expectations of a desired model espoused in Chapter 1, where the problem required the creation of a model that could be used with an autonomous vehicle in new and previously unseen areas. In this chapter, we showed that the proposed framework corresponds to the latter requirement. Consider the results presented for the experimental case on inferring with no training trajectories. This case corresponds to an autonomous vehicle equipped with a sensor array arriving at an entirely new urban environment. The car must then begin to perform predictions on the pedestrians present in the scene without having the luxury of waiting to gather many partial trajectories to train itself in this environment and then perform inference. In demonstrating that the proposed method preforms better than the trained competing model, we showcase the benefit of utilising the proposed method in such a way as to give credence to the "Look-Once Principle". The issue of utilising the framework on an autonomous vehicle with the corresponding sensor systems and the problems inherent to this perception will be dealt in the next chapter.

8 Experimental Results - Egoperception

In the previous chapter, we presented the results of our experiments with the Prior Knowledge Framework in a built urban area with pedestrians and other shared space users being observed from an observer placed externally from an overhang. While this allowed us to test our framework's different capabilities, the ultimate utilisation of this framework is expected to be deployed on an autonomous vehicle. Given such an application, the framework must be able to work with a myriad of sensors, perhaps without visual cues in some instances. Another point to note is that the utilisation of the framework is predicated on an ego centric perception *i.e.,* from the perspective of the autonomous vehicle. Secondly, pedestrian behaviour depends entirely on the built environment they move in. For example, behaviour inside a parking lot differs from that of an open, well defined street in an urban center. Given these constraints, it becomes imperative to test our proposed framework in different urban areas. The results in the previous chapter expounded on the results of our experiments in a well defined, built area. On the contrary, in this chapter, we concentrate on the testing the Prior Knowledge Framework in a private, semi-built area (*in this context, a parking lot*) with limited and ambiguous (*no camera images*) perception.

We start by describing the observation platform and the data captured over time followed by the implementation details of the framework for this application. This is then followed by experimental results, both qualitative and quantitative, leading to a discussion on the obtained results.

8.1 Dataset Description

This dataset was created by capturing data from a Renault Zoe that is used as an experimental platform, described in Section 8.1. The data is captured within the parking lot of the Inria-Grenoble premises. The car is parked facing the main entry points of the building containing two entry/exit points. Not only are cars parked here, there is also a significant point of interest in the form of a bicycle stand which sees many footfalls throughout the day. In real terms, the observed environment covers an area of 60 x 50 sq. meters within the private parking area of the premises.

Pedestrians arrive on the scene from the main entry gate of the parking lot (out of view in the dataset) either by foot or vehicles (car or bicycle), park and enter the building via

the entry points in front of the car. During the time of exit, this procedure is reversed. The parking lot also contains some amount of foliage between car parking areas, which are visually marked out.

The instrumented platform (Fig. 8.2) was placed early in the morning in this area to capture pedestrian behaviour. All actions contained within this dataset were captured exclusively with the available LiDAR on the platform. The LiDAR image as captured by the platform is seen in Fig. 8.1.

The scenario presented by this parking lot dataset is interesting for two reasons -

1. A parking lot is a partially structured environment with specific areas marked out providing their own functions. There are natural POIs in the environment in terms of entrances and exits to buildings and shelters. Thus, pedestrians behave comparably but differently to that of a street (*e.g.*, walking towards a destination in their view but with no mandated walking areas to guide their paths).

2. The parking lot allows for observation of long term evolution of pedestrian trajectories. The wide open areas in the observation environment imply that pedestrians have to walk through for some tens of meters from their vehicles to the entrance/exit to the building. Using only LiDAR data aids in capturing this full trajectory.

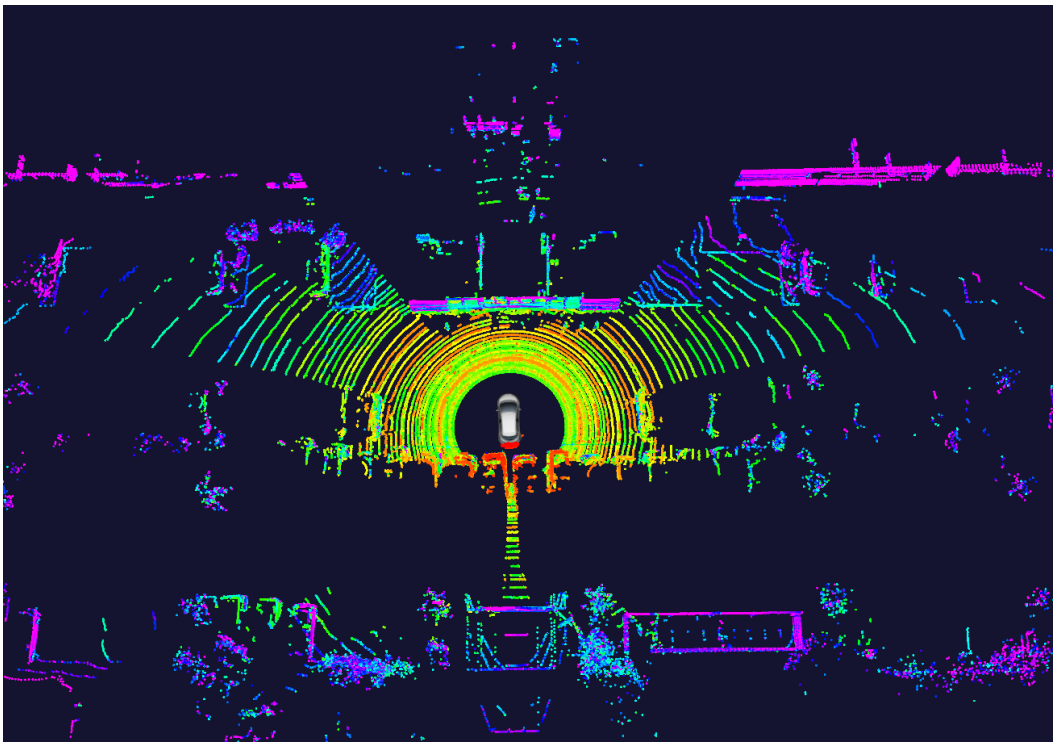


Fig. 8.1.: LiDAR image of the environment under observation, captured by a velodyne, at the Inria parking lot.

Fig. 8.2.: Renault Zoe instrumented platform used for capturing data used for evaluating the proposed algorithms. On the roof of this car is the velodyne used to record the dataset.

Experimental Platform

In order to capture and validate our models on pedestrian behaviour, an in-house experimental platform was used. The base of this platform is a Renault Zoe on which are equipped a 64 layer *Velodyne* LiDAR, 4 *IBEO LUX* laser range finders mounted with one mounted in each cardinal direction on the car, 1 Point Grey Bumblebee stereo camera pointing front mounted inside the car platform, 1 mono back camera and an *XSENSE* IMU GPS receiver to localise the platform on a map. The platform can be seen in Fig. 8.2.

Extraction of trajectories

The extraction of trajectories involved transforming the LiDAR data into 2-D images by first cropping them to a region of interest. These 2-D images were then manually annotated for "pedestrian" and "car" classes using the VATIC tool (Vondrick et al., 2012) for labelling moving objects in a video. While there are methods available to automatically extract semantic information from 3-D point clouds (Zhou and Tuzel, 2017; Qi et al., 2017; Garcia-Garcia et al., 2016), they require much more information than what is available from the instrumented platform.

The feet of the pedestrians were tracked over the course of their movement and their velocities were estimated using a Kalman filter as part of post-processing these observed trajectories. Annotations were made manually on their destinations when they entered one of the possible POIs which was when their trajectories terminated.

Semantics and Geometry

The identification of the different semantics in the scene is relatively straight forward. There are buildings and shelters that can be identified, then there are the parked cars. There are also dynamic obstacles in the form of entering and exiting cars. These can be easily identified and tracked via the utilisation of methodologies like the ones presented in Rummelhard et al. (2015) which provides us with both a bounding box identifying the moving object as well as estimations of its velocity over time. As can be seen in Fig. 8.1, there are some areas that are occluded to the LiDAR. These areas can be semantically considered "unknown" since they cannot be directly observed. Finally, everything else that remains in the observed environment is considered traversable for pedestrians.

On the topic of geometry and the attainment of geometric data, since we are using a LiDAR, it becomes trivial to measure the dimensions of any object, if it is directly observed. Spatial maps depicting the marked parking spots in the parking lot can be obtained though a map service like Open Street Maps[1].

All pedestrian trajectories extracted from this dataset are presented in Fig. 8.3. In contrast to the trajectories presented in Fig. 7.3, this scenario presents a semi-structured environment. Thus, the trajectories are dispersed over the entire observed area with the entrances of the building structure standing out as the termination points to a plurality of the observed trajectories. An interesting fact to note on analysing the trajectories is that some trajectories follow specific, if unmarked, tracks. This example stands as a specimen of "Natural Movement" (Gibson, 1979) where pedestrians move in the direction that interests them the most – in this case, the entrance to the building structure – which is directly in their line of sight.

8.2 Experimental Results

The experiments conducted on this dataset differ slightly from the ones presented in the Chapter 7. Within the semi-structured environment of the parking lot, there are no legal and illegal trajectories. Instead, trajectories wind between obstacles in the direction of Point of Interests. Thus, the experiment described here ultimately is concerned with the accuracy of prediction of pedestrian motion in a scene devoid of interference to their trajectory from other moving objects.

8.2.1 Implementation Details

The architecture of the framework remains the same as the one presented in Section 7.1. The methods used to identify the features in the observed environment and to estimate pedestrian tracks have been mentioned in the preceding sections. Here, we concentrate on the

[1] www.openstreetmap.org

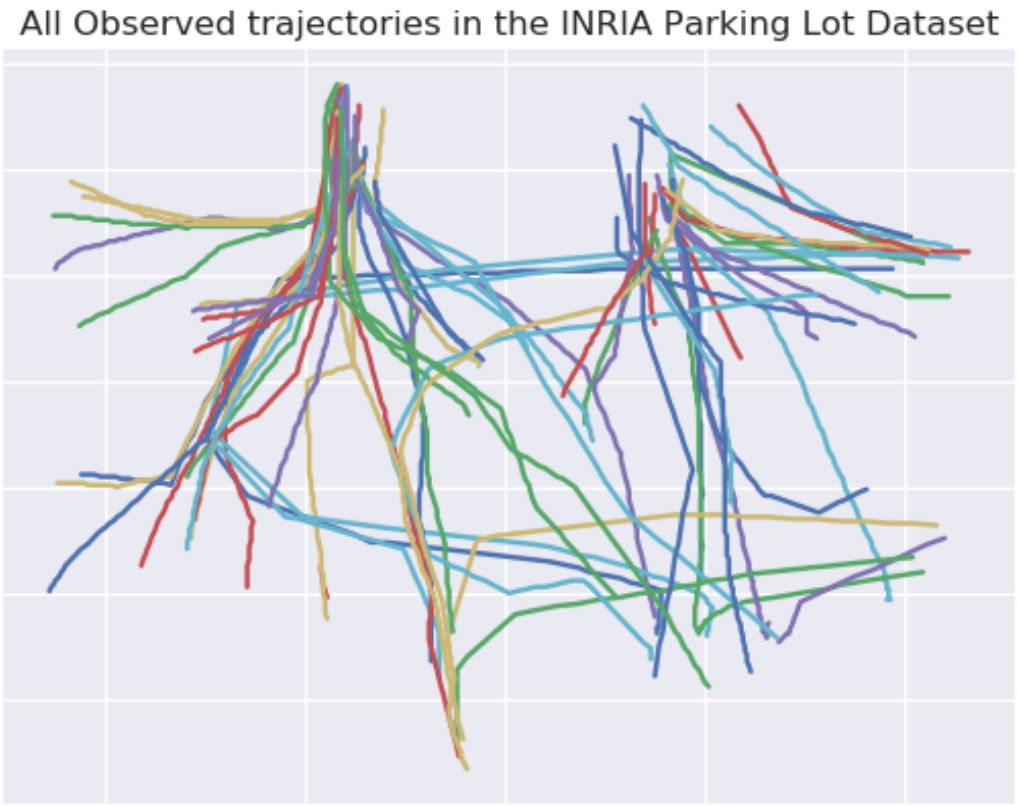


Fig. 8.3.: Tracks of pedestrians extracted from the Inria parking lot dataset observed from an on-board sensing system on the experimental platform.

implementation details of the framework. The velodyne on top of the car returns distance measures in meters. The POIs which are the entrances to the building are considered to be more attractive than the others. Since there is no identifiable "Road" feature in the scene, this feature is discarded. Instead, short barriers are considered edges in the environment. The insertion distance τ is assigned a value of $2.5m$. This value was so chosen to balance between an explosion of nodes in the large environment and a meaningful value to predict a pedestrian's next position.

Similar to the previous chapter, for the competing approach, the value of τ remains equal to the proposed method while π_0 and a_0 are, once again, arbitrarily chosen to be 0.5 and 0.5 respectively.

The utilisation of these parameters results in the generation of the potential cost map and the prior topological map as shown in Fig. 8.5.

8.2.2 Qualitative Results

The qualitative metric for measuring the results remains the same as the one proposed in Chapter 7. Similarly, we compare the results against the model proposed in Pérez-Hurtado

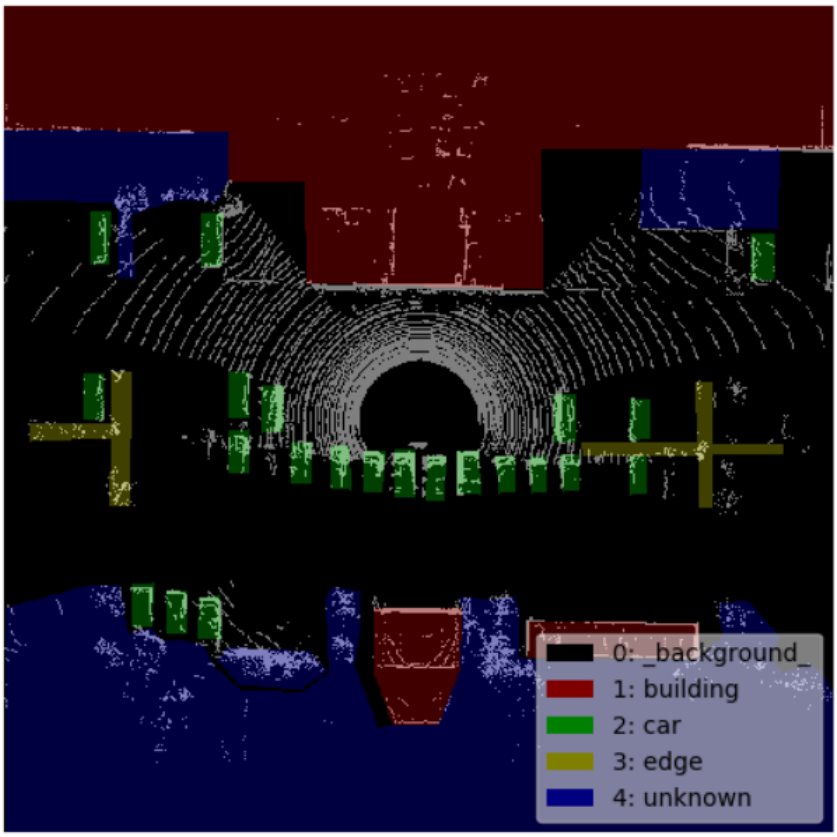
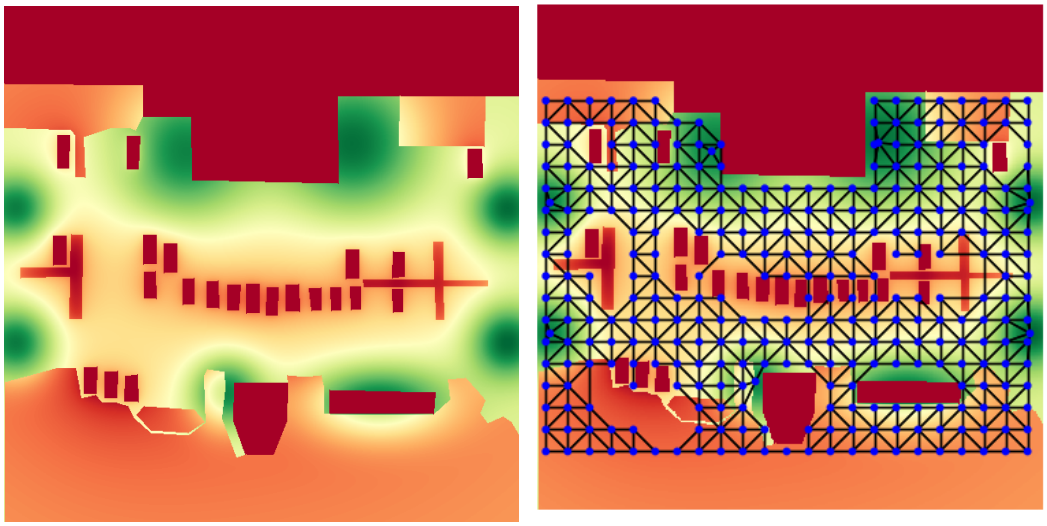


Fig. 8.4.: Labelled image of the Inria parking lot used for the generation of the potential cost map.

et al. (2015). To reiterate, the quality of the achieved result depends on the "coverage" of the graphical model of the observed area. Hence, higher the coverage, better the quality of prediction.

To portray the premise of the qualitative results, consider the testing trajectories depicted in Fig. 8.6. Of the 14 trajectories chosen for testing the performance of the framework from the dataset, three trajectories are displayed. Here, all three trajectories start in the vicinity of the same POI (*bicycle parking shed*) while two terminate at one of the entrances to the building while the remaining trajectory terminates at the other entrance of the building. In our qualitative evaluation, the predictive framework must be able to accommodate these trajectories within the graphical model *i.e.,* be able to capture this behaviour.

Consider the specimen trajectories terminating at the left entrance of the building in Fig. 8.6. Fig. 8.7b depicts the the evolution of the topological map over 50, 100 and 200 training partial trajectories. On applying the qualitative evaluation criterion *i.e.,* that the topology being able to accommodate the trajectories over the lifetime of the trajectory, it is noticed that this criterion does not hold true. In the cases of 50 and 100 training partial trajectories, a large area corresponding to the left side of the grid of Fig. 8.6 is not covered by the topological maps' evolution. The implication here is that there is no possibility for



(a) Cost Map generated using the labelled image presented in Fig. 8.4. The degree of affordance in the observed scene varies from *green* to *red* with *red* depicting areas that are harder to reach (including obstacles) and *green*, the contrary.

(b) The Prior Topological Map generated based on the potential cost map depicted adjacently. The nodes and the edges depict reachable areas in the observed environment.

Fig. 8.5.: Potential Cost Map and Prior Topological Map for the observed environment in the Inria Parking Lot.

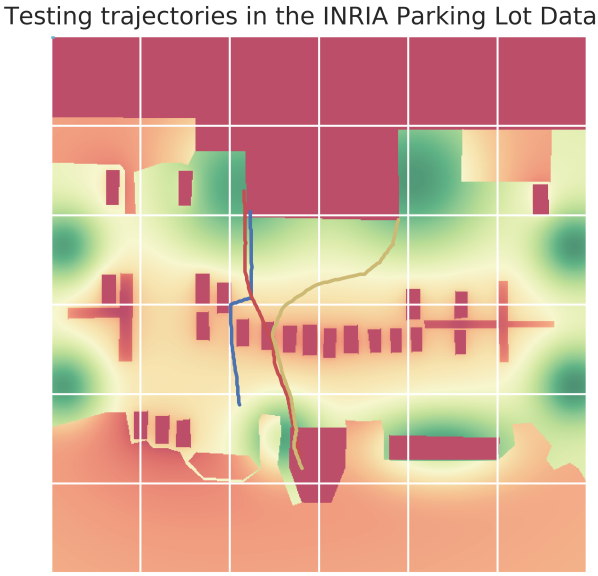
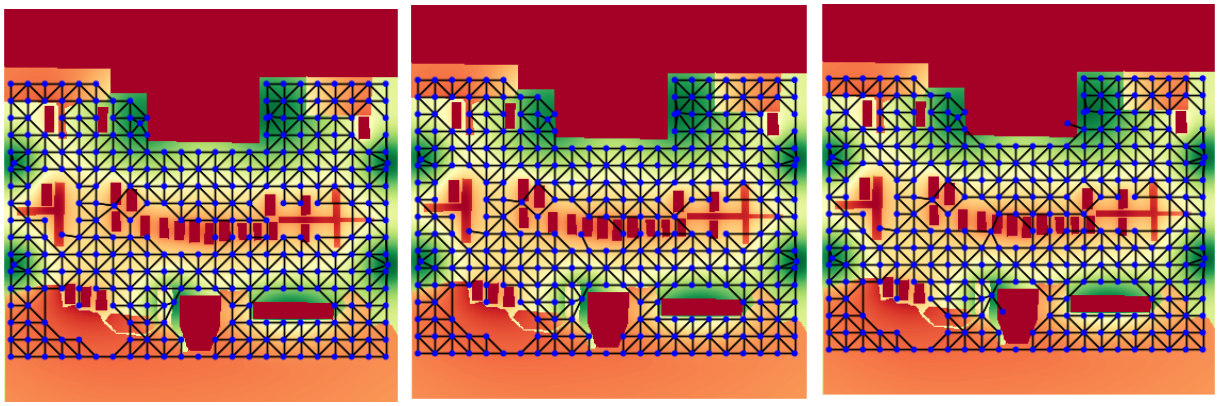
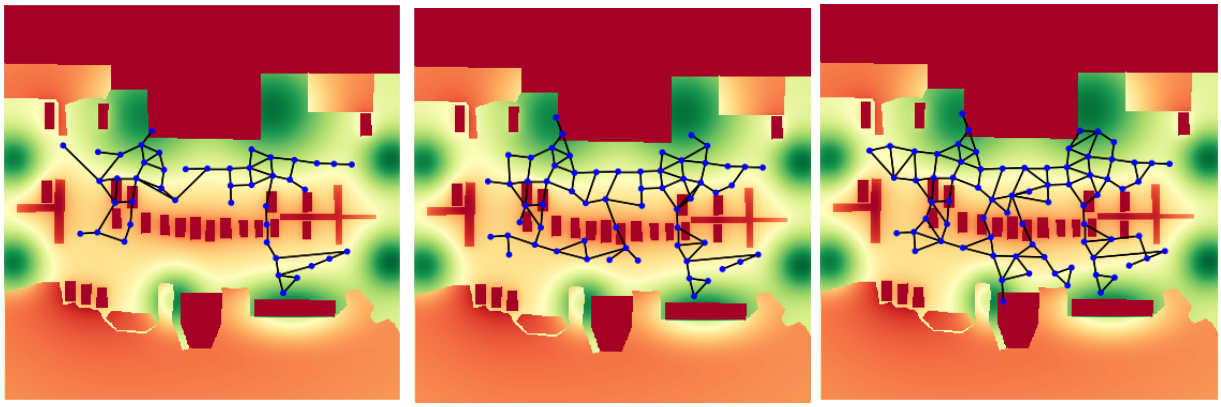


Fig. 8.6.: Specimen testing trajectories chosen from the Inria Parking Lot Dataset to describe the qualitative and quantitative results. The red trajectory passing through the obstacle is an error of the VATIC tracking system that linearly extrapolates trajectories between two observed positions, serving, once again, as a warning as to the importance of the perception layer for the framework.



(a) Evolution of the topological map under the proposed method for 50, 100 and 200 training partial trajectories. This evolution takes into account the environmental context during the update phase.



(b) Evolution of the topological map under method proposed in Pérez-Hurtado et al. (2015) for 50, 100 and 200 training partial trajectories. The evolution of the map does not take into account the environment.

Fig. 8.7.: Qualitative comparison in the evolution of the topological map between the proposed approach and the competing approach in the Inria Parking Lot.

these topological maps to assign an observation from the trajectory to a node in the map thereby failing to capture their motion. On the other hand, with the topological map updates after 200 partial trajectories, it is observed that an edge exists close to the POI, providing a tenuous link from the origin of the trajectories to the termination points over their lifetimes.

On the contrary, consider the states of the topological map derived from the proposed method in Fig. 8.5b with only prior knowledge derived from the environment and in Fig. 8.7a over the course of 50, 100 and 200 training partial trajectories. It can be observed that at each of these topological maps, there always exists nodes and edges to accommodate observations of the specimen trajectories over their entire lifetime thus meeting the qualitative evaluation criteria.

8.2.3 Quantitative Results

The quantitative results deal with the accuracy of predicting the specimen trajectories over a specific time horizon. We continue using the same metric to measure accuracy as in Chapter 7.

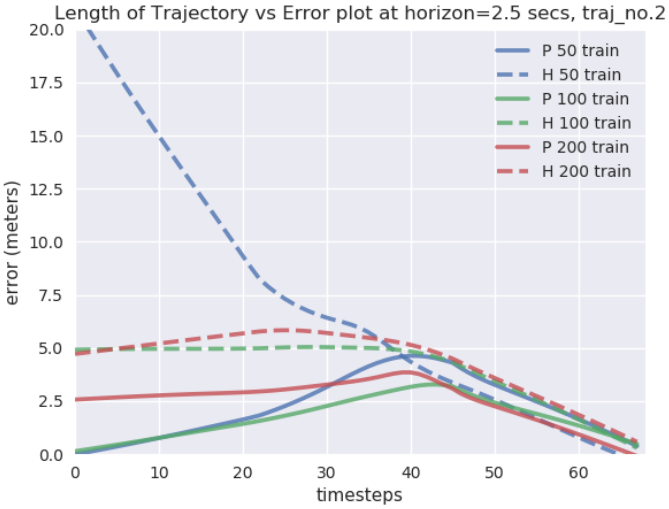
The GHMM models via the proposed approach and the competing approach are trained on the same partial trajectories that resulted in the topological maps described in Fig. 8.7. At the end of the training epochs of 50, 100 and 200 partial trajectories, the trained GHMM models are tested on each trajectory of the testing trajectory set. For every timestep of each trajectory, pedestrian positions at a horizon of 2.5 seconds are predicted. These positions are then compared to the ground truth to establish the Euclidean distance error between the prediction and the ground truth.

A second test is conducted in a similar fashion in testing the accuracy of our proposed model with no training trajectories as compared to a fully trained[2] GHMM model as proposed in Pérez-Hurtado et al. (2015). Prediction is performed for all the available testing trajectories as above.

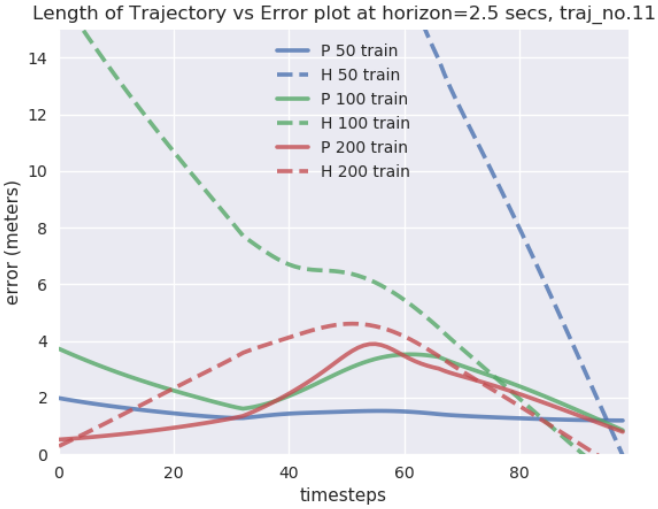
To explain the quantitative results, we use, once again, the same specimen trajectories as before. The qualitative results for the first experiment with prediction errors tested at different training epochs are as shown in Fig. 8.8 with the figures 8.8a, 8.8b and 8.8c representing the errors for the three specimen trajectories.

In the previous section, the problem with the accommodation of the observations of the testing trajectories by the underlying graphical model was touched upon. The quantitative results presented here support the qualitative evaluation criteria presented previously. Consider the error prediction graph for the first specimen trajectory starting at the bottom of the observed environment and terminating in the left entrance of the building premises (depicted by the blue line in Fig. 8.6). This trajectory will be tested for the topologies and hence the underlying GHMM graphical models for all the cases shown in Fig. 8.7. For the topologies presented in Fig. 8.7a at each training epoch, there exists path in terms of nodes and edges in the topology and thus as states and transitions in the corresponding GHMM to be able to predict with sufficient accuracy the future positions of this trajectory given the observations and beliefs around each state of the model over the lifetime of the trajectory. On the other hand, in the model that is compared against, presented in Fig. 8.7b, we observe a marked difference. For the 50^{th} partial training trajectory epoch, the closest state in the model with respect to the tested trajectory is quite far. This leads to a large prediction error in the beginning of the trajectory since there are states close by that can realistically infer the presence of a pedestrian in the specified time horizon. Over the passage of time, with the same epoch, we see that the error decreases since now, the trajectory is at a point where the graphical model can accommodate the rest of the trajectory. Similarly, for the other two

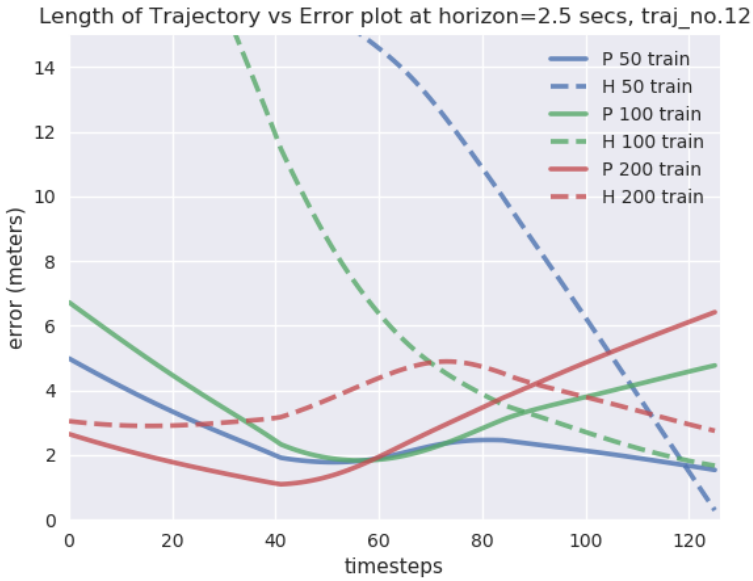
[2] a fully trained model in our context is a GHMM trained with all available partial trajectories



(a) Error comparison graph for the first specimen trajectory terminating in the left entrance to the premises under different epochs of training.



(b) Error comparison graph for the specimen trajectory terminating in the right entrance to the premises under different epochs of training.



(c) Error comparison graph for the second specimen trajectory terminating in the left entrance to the premises under different epochs of training.

Fig. 8.8.: Graphs depicting the prediction errors between the proposed method and that of competing approach. Each graph shows the evolution of errors over the lifetime of the trajectory under different epochs of training. These epochs are 50, 100 and 200 partial training trajectories. Solid and dashed lines of the same colour depict the proposed method and the compared method, respectively for the specimen testing trajectories. The graphs depict error in meters on the y-axis and the x-axis is composed of the length of the trajectories in timesteps.

	Number of Partial Training Trajectories			
	50	100	200	0 - Full
Testing Trajectories	7.07×10^{-4}	2.6×10^{-3}	4.1×10^{-4}	2.4×10^{-4}

Tab. 8.1.: Comparison of prediction accuracy at horizon=3 secs at varying training levels between the proposed work and the competing approach. Lower is better for the values.

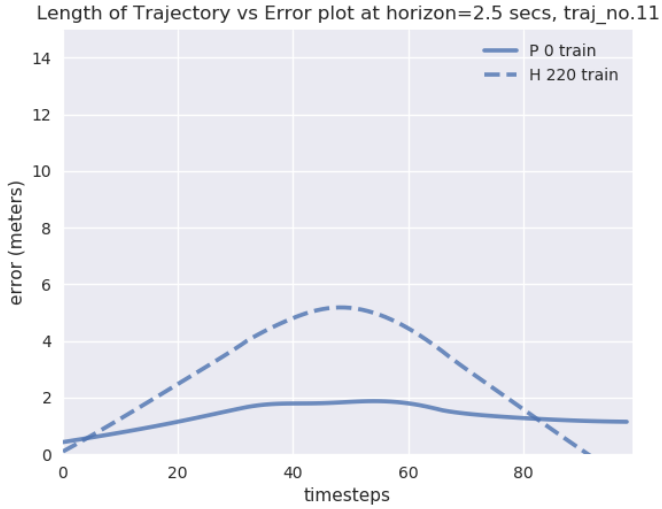
epochs *i.e.,* for 100 and 200^{th} training epochs, there are more nodes and edges covering the observed area, correspondingly reducing the prediction error for the compared model. On the other hand, for the proposed model, there is no such abrupt, significant change with the change only making the system more accurate in its prediction.

Similar inferences can be made for the two other testing trajectories. At low training epochs, the model compared against performs worse than the proposed model while the proposed model continues to increase the prediction accuracy over training epochs. An exception to this is the prediction error for the third trajectory presented in Fig. 8.8c. Here, at higher training epochs, near the end of the lifetime of the trajectory, the prediction accuracy decreases. This can be explained by the fact that there are numerous states and the belief in pedestrian position is distributed over many states leading to a slight increase in the inaccuracy.

The results of the second experiment *i.e.,* that of testing the proposed model with no training against the fully trained compared model can be seen in Fig. 8.9. The results follow in a similar fashion as before, with the proposed model performing as well as or better than the compared model. The reason remains the same as before, since there are an insufficient number of training trajectories as to cover all possible pedestrian paths in the environment and to train the underlying HMM as to represent trajectories accurately. The untrained proposed model performs well since each state is instantiated with a gaussian probability, albeit circular, associated with it when created. This property makes it such that there is a belief over the states even with no training trajectories, contrary to the method used in the compared approach, tying in, once again, to the qualitative evaluation criterion of having good coverage of the observed environment.

As in the previous chapter dealing with exo perception, a *t*-test is conducted to ascertain whether the prediction results are a significant improvement over the compared method. The result can be seen in Table 8.1. As can be seen in the table, all the *p*-values are lower than our significance level of 0.05. Thus, for ego perception, as for exo perception, the proposed framework performs at a significantly better accuracy.

(a) Error comparison graph for the first specimen trajectory terminating in the left entrance to the premises under no training to full training.



(b) Error comparison graph for the specimen trajectory terminating in the right entrance to the premises under no training to full training.

(c) Error comparison graph for the second specimen trajectory terminating in the left entrance to the premises under no training to full training.

Fig. 8.9.: Graphs depicting the prediction errors between the proposed method with no training data *vs.* the competing method after full training. Each graph follows the procedure of the previously described environment with the proposed model with no training data and the model compared against trained with all available training partial trajectories. Solid and dashed lines represent the proposed method and the compared method, respectively for the specimen testing trajectories. The graphs depict error in meters on the y-axis and the x-axis is composed of the length of the trajectories in timesteps.

8.3 Discussion

In this chapter, we presented the experimental results of the proposed approach using ego-perception from an autonomous vehicle on real data captured in a semi-structured, closed parking lot. To perform this, we utilised three specimen trajectories, starting at one POI and terminating at two other POIs, to illustrate the qualitative and quantitative results of our model with the results presented in Section 8.2. We demonstrated that our proposed method performs better than the compared approach at different epochs of training, thereby proving that it is possible to utilise this method from the perspective of an autonomous vehicle. Yet, some limitations of the proposed approach can be noticed in the results.

The first limitation, is once again the problem of good perception, especially for tracking trajectories reliably. For the experiments conducted in this chapter, as has already been mentioned, only LiDAR data was used to identify and track pedestrians. In such a setup, there are chances of pedestrian trajectories, while tracking, are extrapolated through areas where they may not have passed through originally leading to strange topological updates.

A second limitation can be noticed in the quantitative results. Here, the accuracy of prediction is based on the errors defined by the Euclidean distance between the predicted position and the ground truth. This prediction is centered on the position of the nodes in the topological map since these nodes act in discretising the observed space with the *Insertion Distance* parameter τ limiting the distance between the insertion of two nodes. In the implementation of the experiment conducted in this chapter, τ was found to be $\approx 2.5\ mts$ when translated to real world co-ordinates. Accounting for this and the point mentioned above, accuracies can only be measured in terms of τ differences. This implies that if the prediction error is 4.5 meters, then the prediction is only off by one node on the topological map which, while from the perspective of the model is not bad, could have real world consequences. One method of solving this limitation could be to reduce τ thereby increasing the discretisation but this brings us to another limitation. Here, with increasing nodes with higher discretisation, there is a need for more training data to predict trajectories more accurately. Increasing the number of nodes proportionally increases the number of states and thereby, with the belief maintained over each state, can lead to wrong predictions when faced with few or no training trajectories. Finding the optimal value of τ to balance accuracy and complexity of the model remains an open question.

Part IV

Conclusion

9 Conclusions and Perspectives

9.1 Summary

This thesis was carried out under the ambit of the ANR VALET project that envisages the operation of a platoon of autonomous vehicles whose leader is a manually driven vehicle. Thus, much of the work was aimed at rapidly modelling new environments, setting the stage for predicting pedestrian trajectories from the perspective of an autonomous vehicle. Chapter 1 introduced the ideas that drove the development of the Prior Knowledge framework, with an emphasis on the "Look-Once" principle requiring the autonomous vehicle to be able to model and begin predicting pedestrian trajectories as soon as it entered a new area.

Chapter 2 began with a short review of the state of the art of pedestrian prediction methods that exist in literature. In this chapter, we presented and discussed the various approaches from those using simple physics-based models for prediction with no learning to those heavily dependent on data. We also presented object centric approaches that depended on the observation of corporeal cues of the pedestrians and environment-centric approaches that emphasised on the effects of the surrounding environment on the pedestrian. We concluded with a discussion on interaction aware approaches where other shared space users have a tangible effect on the pedestrian whose intention is to be predicted.

Chapter 3 and Chapter 4 presented a short introduction on the probabilistic models used in this thesis. The former emphasised heavily on the usage and shortcomings of Hidden Markov Models in their utilisation for prediction tasks while the latter began with an introduction into an extension of HMMs called the Growing Hidden Markov Model (GHMM). In that chapter, we then presented a detailed look at the inner workings of the GHMMs and the algorithms used, specifically those utilised to learn the parameters and the topologies in the observed space. We concluded by detailing the limitations of the original implementation of the GHMM method with a direction on how to solve them.

Chapter 5 presented our first contribution in the thesis. This contribution was an algorithm that converted the various features in the observed environment and from them, generated a potential cost map. We also introduced in this chapter, the idea of Point of Interest (POI)s derived from sociological ideas of *Natural Vision* and *Natural Movement* that could be used to translate pedestrian movement within a built, structured urban area into quantifiable costs that aid in understanding, predicting and generating natural pedestrian behaviour.

Chapter 6 described our second contribution, the extension to the GHMM method. We presented in this chapter how a generated cost map can aid in creating a better topological map that can accommodate atypical trajectories and initialise a better underlying graphical model that requires fewer training epochs to preform prediction tasks under different conditions. We began by juxtaposing the existing state of the art in GHMMs for pedestrian prediction with its limitations. Subsequently, we proposed our algorithm targeting these identified limitations. Finally, we presented the third contribution of the thesis - the utilisation of the GHMM method to predict pedestrian positions within a dynamic scene where the various moving objects affect the trajectories a pedestrian can take in the environment under observation.

Chapter 7 presented, first and foremost, the proposed framework called the Prior Knowledge Framework that conforms to the "Look-Once" principle outlined in Chapter 1. This framework is also the culmination of the other contributions described in earlier chapters and contains a detailing of the modules it is comprised of. Next, this chapter contains a description of the dataset used to arrive at the results. The results contained within the chapter are based on the annotations on video where the images captured are from a camera placed at an angle, overlooking a busy cross-walk in an urban center. Here, we compared our results, obtained by parsing all the required information through the proposed framework, qualitatively and quantitatively against the GHMM model proposed in Pérez-Hurtado et al. (2015) at different training epochs, trained on partial trajectories and tested against "legal" and "illegal" behaviour of pedestrians in this built area. Here, we succeeded in demonstrating that our proposed method performs better under different training epochs against the compared model. A significant demonstration was that when tested under the condition of no training of our proposed model *vs.* a fully trained competing model resulting in equivalent or better performance in comparison. This makes the proposed approach a good candidate for use in new and previously unobserved areas where there is need for rapidly adapting to the environment to predict pedestrian trajectories before any training observations are possible. In continuation, we also tested our framework by including dynamic objects in the scene so as predict trajectories under natural conditions and compared against results on the same trajectory obtained via our framework with no training. A significant constraint under this class of experimentation was that the inference was performed with the state of the environment frozen at one point over the lifetime of the trajectory whose positions were being predicted. Within these constraints, the framework performed prediction admirably, under certain conditions while failing under others. These limitations were discussed at the conclusion of the chapter.

Chapter 8 produced results from a scene that could be juxtaposed against the previous chapter. While the previous chapter dealt with testing the framework on data derived from a static camera with a bird's eye view of the observed area, this chapter dealt with data obtained from a LiDAR placed on the top of a moving, testing platform which is more in

line with the perception associated with the real-world operation of an autonomous vehicle. These observations are recorded within a parking lot of a building. Contrary to the perception from the previous chapter, this perspective of observation leads to occlusions in many parts of the environment. Another factor in utilising this data is the fact that a parking lot is a semi-structured environment, changing the motion patterns of pedestrians within it. Testing our proposed framework with these factors becomes paramount, especially if it is to be deployed in the real world. We presented the details of this dataset and obtained results in line with the previous chapter. We concluded with a discussion on some of the limitations of the framework based on the results.

9.2 Concluding Remarks and Perspectives

During the course of this work, the following conclusions were reached:

- Sociological principles of attraction and repulsion from different features present within an urban scene such as crosswalks, sidewalks, road, edges, POIs *etc.,* can be captured and quantified using principles potential fields to generate what we call "Potential Cost Maps". This confluence of sociological ideas, such as POIs acting as destinations for pedestrians, and the quantification of these attractions and repulsions in the scene leads to a quicker approach in modelling the environment to be able to generate more natural pedestrian trajectories, as was demonstrated in Section 5.4.

- Based on the results obtained in chapters 7 and 8, it was demonstrated that the proposed framework can be utilised in a variety of situations - using visual input; with non-visual, partially occluded input; in structured and unstructured environments; in static scenes and in dynamic scenes in the context of typical and atypical trajectories performed by pedestrians.

While these advantages depict the proposed framework to be apt for utilisation on autonomous vehicles, there are also some limitations to its utilisation.

- To begin, studies like Feng et al. (2013) and Cœugnet et al. (2019) show that that while the forward velocity of every moving object is repulsive in nature, there exists an attractive area preceding them. The same is applicable to crowds of pedestrians trying to cross the road on crosswalks as well as performing illegal crossings. This modelisation, while not captured in the current work, could lead to a better understanding of pedestrian behaviours and thereby aid in performing better predictions.

- In chapters 7 and 8, it was noticed that the quality of prediction suffered at very high training data or at the point where there were too many states and not enough training

data. As was mentioned previously, the balancing between these two cases remains an open problem.

- The framework, in its current iteration, has been observed to have been useful for short to medium term prediction in terms of accuracy. Such a design has been chosen so as to be used on an autonomous vehicle to perform risk analyses and take decisions in novel environments while driving. With such requirements, there is little purpose in designing accurate prediction systems for longer horizons, especially if such designs are offset by an increase in the number of observations or training data required. Nevertheless, the extension of the framework for prediction at longer horizons remains an open problem with some directions provided in the following section.

- During the conduction of experiments, especially those in chapter 7 concerning the experiments with the inclusion of moving objects, the procedure followed was to freeze the environment at a specific timestep over the lifetime of the testing trajectory and continue the prediction process by feeding the framework with this state of the environment and continuing from that timestep onwards for a time horizon. During the implementation of this framework on a real autonomous platform, it is envisaged that the potential cost map generation as well as the generation of the topological map and the underlying GHMM is regenerated for ever time step. This, albeit costly, process for short horizons can perform at a better accuracy over the lifetime of the tracked pedestrian.

- A second limitation noticed with the inclusion of moving objects in chapter 7 concerns the lack of interaction between the different shared space users in the environment. Consequently, the current state of the framework does not take into consideration that this interaction could lead to atypical behaviours during inference. The inclusion of this interaction remains a problem that could be questioned in a future work. A solution to this problem could be reached by the introduction of the social force model at inference time as can be found in the work of Elfring et al. (2014).

- Lastly, the framework has been tested on single pedestrians in all environments and cases. Using it for prediction with crowds or multiple people remains a challenge, since such a scenario is faced by vehicles, both human and autonomous, everyday. While the inference for multiple people over the same topological map seems a trivial extension, that with the incorporation of crowds appears to be a more challenging task. One avenue for finding a solution could be to explore the domain of sociology once more, to find cues on crowd dynamics in urban areas.

9.3 Future Work and Possible Extensions

While the current work was aimed at increasing the Situational Awareness of a single autonomous vehicle, the premise of the VALET project creates opportunities for multiple extensions. These extensions can broadly be grouped as high level extensions *i.e.,* dealing with increasing the SA at an organised group of AVs, for instance, at the platoon level or low level extensions, dealing with the modification of the proposed framework itself.

High Level Extensions:

- **Creating Better Maps** A platoon of autonomous vehicles can be disparate with each of the constituent vehicles having asymmetrical perception capabilities, compromising the quality of behaviour prediction of those vehicles which are not as well instrumented as others in the same platoon. In such scenarios, a shared map generated using the perception capabilities of the more instrumented vehicles can solve problems that could lead to a loss of SA. For example, the lead vehicle could generate the cost map for the area under observation, identify and track pedestrians while the other vehicles in the platoon could learn their behaviour and infer their positions better in relation to the current state of the platoon. Such a mental map reduces the re-learning and prediction overhead for each of the vehicles while moving in the same environment. Similarly, these maps could be shared with V2X (Vehicle-to-infrastructure) kiosks placed around these urban areas which can, subsequently, share this information with other AVs on demand.

- **Avoiding the handcrafting of POI weights** The current modelling approach requires the description of the relative importance of the different POIs in the observed scene, *i.e.,* providing the weights of each of the POIs present. Within the ambit of the VALET project, the platoon is expected to traverse pre-determined corridors in the urban area which can be exploited to avoid the handcrafting of these weights. Pedestrian trajectories observed for inference can, post facto, be run through Inverse Reinforcement Learning algorithms with regard to their destinations and to different POIs to learn their relative weights over time leading to the creation of better maps.

Low Level Extensions:

- **Integration of motion models** During the prediction phase of the framework, motion models pertaining to pedestrian motion can be included to get more accurate predictions during inference. Motion models such as Switching Linear Dynamic Systems (SLDS) combined with the underlying graphical model should pave the way for better predictions, especially when predicting pedestrian intentions with other shared space

users where people are observed to change their velocities and directions abruptly to avoid collisions *etc.*

- **Addition of other shared space users** The prediction framework presented in this thesis deals exclusively with pedestrians in urban areas. However, this work can be extended to encompass the behaviours of other shared space users like cars and bi-cyclists in urban spaces. This extension could be a straight forward application of Dynamic Bayesian Networks placed over the existing pedestrian prediction graphical model presented in this work to model their interactions.

- **Introduction of Corporal Cues** The presented framework takes only the velocity and heading of pedestrians into account during the inference phase to predict their positions. Prediction of pedestrian behaviour can be greatly improved on taking into consideration some of the other cues used by human drivers to decide whether a pedestrian will cross or not. This extension could incorporate the cues such as shoulder and body poses (Ghori et al., 2018), distance from the edge of the side-walk (Völz et al., 2016) or attention being paid to the driver of the car (Rasouli et al., 2017).

- **Utilisation of Neural Network Architectures** Current state of the art techniques are moving away from traditional learning methodologies and adapting data-hungry neural networks to learn behaviours and perform inference. A special implementation of the Recurrent Neural Network called the Long-Short Term Memory (LSTM) networks have been used successfully as replacements in domains such as Speech recognition (Graves and Jaitly, 2014) and handwriting sequence recognition (Graves, 2013) which have, traditionally, been applications of HMMs. A recent work by Alahi et al. (2016) has proposed an LSTM neural network architecture to perform short term inference of pedestrian behaviour in social conditions. While the cited literature performs well under specific conditions of social interactions with other pedestrians in an open area, it does not specifically aim to solve the problem presented in the proposed framework. While the performance of our framework has not been tested against this LSTM network, future work could focus on the utilisation of such networks for urban scenes requiring interactions with other shared space users to supplement or supplant the underlying graphical model to ensure accuracy of pedestrian predictions in the observed environment.

Over the course of this thesis we set out to solve the challenge of making Autonomous Vehicles perceive the environment like human drivers and to ensure the capture anomalous pedestrian behaviour. In proposing our algorithms and framework that could be used with little to no training data, and a simple inference procedure, we have managed to reduce our dependency on massive amounts of data and all the inherent issues that follow it. While the aspiration of this work is to make Autonomous Vehicles perform better in urban areas, the

spirit in which it was conducted amounts to a hope that this will lead to safer streets where sudden, anomalous behaviour is captured early and appropriate decisions are taken rapidly so as to avoid needless, and tragic, accidents. As a parting thought, such behaviours by autonomous vehicles could certainly lead to their quicker acceptance, opening up avenues for the betterment of human society.

Bibliography

Agamennoni, Gabriel, Juan I Nieto, and Eduardo M Nebot (2012). "Estimation of multivehicle dynamics by considering contextual information". In: *IEEE Transactions on Robotics* 28.4, pp. 855–870 (cit. on p. 24).

Alahi, Alexandre, Kratarth Goel, Vignesh Ramanathan, et al. (2016). "Social lstm: Human trajectory prediction in crowded spaces". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 961–971 (cit. on pp. 25, 142).

Badrinarayanan, Vijay, Ankur Handa, and Roberto Cipolla (2015). "Segnet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling". In: *arXiv preprint arXiv:1505.07293* (cit. on p. 109).

Ballan, Lamberto, Francesco Castaldo, Alexandre Alahi, Francesco Palmieri, and Silvio Savarese (2016). "Knowledge transfer for scene-specific motion prediction". In: *European Conference on Computer Vision*. Springer, pp. 697–713 (cit. on p. 8).

Bandyopadhyay, Tirthankar, Chong Zhuang Jie, David Hsu, et al. (2013). "Intention-Aware Pedestrian Avoidance". In: *The 13th International Symposium on Experimental Robotics*, pp. 963–977 (cit. on pp. 23, 55).

Barker, Roger Garlock and Herbert Fletcher Wright (1954). *Midwest and its children: The psychological ecology of an American town*. Row, Peterson Evanston, IL (cit. on p. 55).

Batkovic, Ivo, Mario Zanonl, Nils Lubbe, and Paolo Falcone (2018). "A Computationally Efficient Model for Pedestrian Motion Prediction". In: *2018 European Control Conference (ECC)*. IEEE, pp. 374–379 (cit. on p. 19).

Baum, Leonard E., Ted Petrie, George Soules, and Norman Weiss (1970). "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains". In: *The Annals of Mathematical Statistics* 41.1, pp. 164–171 (cit. on p. 35).

Bennison, DJ and RL Davies (1977). "The local effects of city centre shopping schemes: a case study". In: *PTRC Summer Annual Meeting Proceedings of Seminar AP*. Vol. 148, pp. 8–19 (cit. on p. 62).

Best, R. A. and J. P. Norton (July 1997). "A new model and efficient tracker for a target with curvilinear motion". In: *IEEE Transactions on Aerospace and Electronic Systems* 33.3, pp. 1030–1037 (cit. on p. 18).

Bonnin, S., T. H. Weisswange, F. Kummert, and J. Schmuedderich (Oct. 2014). "Pedestrian crossing prediction using multiple context-based models". In: *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pp. 378–385 (cit. on p. 22).

Borgers, Aloys and Harry Timmermans (1986). "A model of pedestrian route choice and demand for retail facilities within inner-city shopping areas". In: *Geographical analysis* 18.2, pp. 115–128 (cit. on p. 62).

Brosseau, Marilyne, Sohail Zangenehpour, Nicolas Saunier, and Luis Miranda-Moreno (2013). "The impact of waiting time and other factors on dangerous pedestrian crossings and violations at signalized intersections: A case study in Montreal". In: *Transportation Research Part F: Traffic Psychology and Behaviour* 21, pp. 159–172 (cit. on p. 93).

Brouwer, N., H. Kloeden, and C. Stiller (Nov. 2016). "Comparison and evaluation of pedestrian motion models for vehicle safety systems". In: *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2207–2212 (cit. on p. 17).

Cai, Yizheng, Nando de Freitas, and James J. Little (2006). "Robust Visual Tracking for Multiple Targets". In: *Computer Vision – ECCV 2006*. Ed. by Aleš Leonardis, Horst Bischof, and Axel Pinz. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 107–118 (cit. on p. 18).

Cannon, Molly Boeka (2013). "Exploring the nature of space for human behavior in ordinary structured environments". PhD thesis. University of Nebraska. (cit. on p. 55).

Casson, Ronald W (1983). "Schemata in cognitive anthropology". In: *Annual review of anthropology* 12.1, pp. 429–462 (cit. on p. 6).

Chan, A. D. C. and K. B. Englehart (Jan. 2005). "Continuous myoelectric control for powered prostheses using hidden Markov models". In: *IEEE Transactions on Biomedical Engineering* 52.1, pp. 121–124 (cit. on p. 36).

Chen, Yu Fan, Miao Liu, and Jonathan P How (2016). "Augmented dictionary learning for motion prediction". In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 2527–2534 (cit. on p. 21).

Cœugnet, Stéphanie, Béatrice Cahour, and Sami Kraïem (2019). "A Psycho-Ergonomic Approach of the Street-Crossing Decision-Making: Toward Pedestrians' Interactions with Automated Vehicles". In: *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)*. Ed. by Sebastiano Bagnara, Riccardo Tartaglia, Sara Albolino, Thomas Alexander, and Yushi Fujita. Cham: Springer International Publishing, pp. 132–141 (cit. on p. 139).

Das, Sanghamitra, Charles F Manski, and Mark D Manuszak (2005). "Walk or wait? An empirical analysis of street crossing decisions". In: *Journal of applied econometrics* 20.4, pp. 529–548 (cit. on p. 63).

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). "Maximum Likelihood from Incomplete Data via the EM Algorithm". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 39.1, pp. 1–38 (cit. on pp. 31, 35).

Deo, N. and M. M. Trivedi (Oct. 2017). "Learning and predicting on-road pedestrian behavior around vehicles". In: *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1–6 (cit. on pp. 21, 79).

Elfring, Jos, René Van De Molengraft, and Maarten Steinbuch (2014). "Learning intentions for improved human motion prediction". In: *Robotics and Autonomous Systems* 62.4, pp. 591–602 (cit. on p. 140).

Ellis, D., E. Sommerlade, and I. Reid (Sept. 2009). "Modelling pedestrian trajectory patterns with Gaussian processes". In: *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, pp. 1229–1234 (cit. on p. 21).

Elnagar, A. (July 2001). "Prediction of moving objects in dynamic environments using Kalman filters". In: *Proceedings 2001 IEEE International Symposium on Computational Intelligence in Robotics and Automation (Cat. No.01EX515)*, pp. 414–419 (cit. on p. 18).

Elnagar, Ashraf and Kamal Gupta (1998). "Motion prediction of moving objects based on autoregressive model". In: *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 28.6, pp. 803–810 (cit. on p. 18).

Endsley, Mica R (1995). "Toward a theory of situation awareness in dynamic systems". In: *Human factors* 37.1, pp. 32–64 (cit. on pp. 3, 4).

Feng, Shumin, Ning Ding, Tao Chen, and Hui Zhang (July 2013). "Simulation of pedestrian flow based on cellular automata: A case of pedestrian crossing street at section in China". In: *Physica A: Statistical Mechanics and its Applications* 392, pp. 2847–2859 (cit. on p. 139).

Ferrier-Barbut, E., D. Vaufreydaz, J. David, J. Lussereau, and A. Spalanzani (June 2018). "Personal Space of Autonomous Car's Passengers Sitting in the Driver's Seat". In: *2018 IEEE Intelligent Vehicles Symposium (IV)*, pp. 2022–2029 (cit. on p. 102).

Ganem, Joseph (1998). "A behavioral demonstration of Fermat's principle". In: *The Physics Teacher* 36.2, pp. 76–78 (cit. on p. 60).

Garcia-Garcia, A., F. Gomez-Donoso, J. Garcia-Rodriguez, et al. (July 2016). "PointNet: A 3D Convolutional Neural Network for real-time object class recognition". In: *2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 1578–1584 (cit. on p. 123).

Ghori, O., R. Mackowiak, M. Bautista, et al. (June 2018). "Learning to Forecast Pedestrian Intention from Pose Dynamics". In: *2018 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1277–1284 (cit. on pp. 23, 142).

Gibson, James J (1979). "The ecological approach to visual perception." In: (cit. on pp. 10, 50, 61, 62, 81, 124).

– (1977). "The theory of affordances". In: *Hilldale, USA* 1, p. 2 (cit. on p. 86).

Goldhammer, Michael, Matthias Gerhard, Stefan Zernetsch, Konrad Doll, and Ulrich Brunsmann (2013). "Early prediction of a pedestrian's trajectory at intersections". In: *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*. IEEE, pp. 237–242 (cit. on p. 22).

Graves, Alex (2013). "Generating sequences with recurrent neural networks". In: *arXiv preprint arXiv:1308.0850* (cit. on p. 142).

Graves, Alex and Navdeep Jaitly (2014). "Towards end-to-end speech recognition with recurrent neural networks". In: *International conference on machine learning*, pp. 1764–1772 (cit. on p. 142).

Habibi, Golnaz, Nikita Jaipuria, and Jonathan P How (2018). "Context-aware pedestrian motion prediction in urban intersections". In: *arXiv preprint arXiv:1806.09453* (cit. on p. 21).

Hashimoto, Yoriyoshi, Gu Yanlei, Li-Ta Hsu, and Kamijo Shunsuke (2015). "A probabilistic model for the estimation of pedestrian crossing behavior at signalized intersections". In: *2015 IEEE 18th International Conference on Intelligent Transportation Systems*. IEEE, pp. 1520–1526 (cit. on p. 25).

He, Kaiming, Georgia Gkioxari, Piotr Dollár, and Ross B. Girshick (2017). "Mask R-CNN". In: *CoRR* abs/1703.06870. arXiv: 1703.06870 (cit. on p. 5).

Helbing, Dirk (1993). "Boltzmann-like and Boltzmann-Fokker-Planck equations as a foundation of behavioral models". In: *Physica A: Statistical Mechanics and its Applications* 196.4, pp. 546–573 (cit. on p. 59).

Helbing, Dirk and Peter Molnar (1995). "Social force model for pedestrian dynamics". In: *Physical review E* 51.5, p. 4282 (cit. on pp. 19, 60, 70).

Hillier, Bill, Alan Penn, Julienne Hanson, Tadeusz Grajewski, and Jianming Xu (1993). "Natural movement: or, configuration and attraction in urban pedestrian movement". In: *Environment and Planning B: planning and design* 20.1, pp. 29–66 (cit. on p. 61).

Hine, Julian and John Russell (1993). "Traffic barriers and pedestrian crossing behaviour". In: *Journal of transport Geography* 1.4, pp. 230–239 (cit. on pp. 50, 62).

Hoogendoorn, Serge P and Piet HL Bovy (2004). "Pedestrian route-choice and activity scheduling theory and models". In: *Transportation Research Part B: Methodological* 38.2, pp. 169–190 (cit. on pp. 42, 60).

Hu, Weiming, Xuejuan Xiao, Zhouyu Fu, et al. (2006). "A system for learning statistical motion patterns". In: *IEEE transactions on pattern analysis and machine intelligence* 28.9, pp. 1450–1464 (cit. on p. 20).

Hughes, Roger L. (2002). "A continuum theory for the flow of pedestrians". In: *Transportation Research Part B: Methodological* 36.6, pp. 507–535 (cit. on p. 60).

Jockusch, J. and H. Ritter (1999). "An Instantaneous Topological Mapping Model for Correlated Stimuli". In: *International Joint Conference on Neural Networks, (IJCNN)*. Vol. 1, 529–534 vol.1 (cit. on p. 45).

Joseph, Joshua, Finale Doshi-Velez, Albert S Huang, and Nicholas Roy (2011). "A Bayesian nonparametric approach to modeling motion patterns". In: *Autonomous Robots* 31.4, p. 383 (cit. on p. 21).

Kaempchen, N., K. Weiss, M. Schaefer, and K. C. J. Dietmayer (June 2004). "IMM object tracking for high dynamic driving maneuvers". In: *IEEE Intelligent Vehicles Symposium, 2004*, pp. 825–830 (cit. on p. 18).

Kanda, T., D. F. Glas, M. Shiomi, and N. Hagita (Dec. 2009). "Abstracting People's Trajectories for Social Robots to Proactively Approach Customers". In: *IEEE Transactions on Robotics* 25.6, pp. 1382–1396 (cit. on pp. 18, 57, 58).

Karasev, V., A. Ayvaci, B. Heisele, and S. Soatto (May 2016). "Intent-aware long-term prediction of pedestrian motion". In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2543–2549 (cit. on p. 24).

Kaufman, Leonard and Peter J Rousseeuw (2009). *Finding groups in data: an introduction to cluster analysis*. Vol. 344. John Wiley & Sons (cit. on p. 20).

Kautz, Henry A and James F Allen (1986). "Generalized Plan Recognition." In: *AAAI*. Vol. 86. 3237, p. 5 (cit. on p. 55).

Keller, C. G. and D. M. Gavrila (Apr. 2014). “Will the Pedestrian Cross? A Study on Pedestrian Path Prediction”. In: *IEEE Transactions on Intelligent Transportation Systems* 15.2, pp. 494–506 (cit. on p. 19).

Khatib, O. (Mar. 1985). “Real-time obstacle avoidance for manipulators and mobile robots”. In: *Proceedings. 1985 IEEE International Conference on Robotics and Automation*. Vol. 2, pp. 500–505 (cit. on p. 64).

Khreich, Wael, Eric Granger, Ali Miri, and Robert Sabourin (2010). “On the memory complexity of the forward–backward algorithm”. In: *Pattern Recognition Letters* 31.2, pp. 91–99 (cit. on p. 37).

Kitani, Kris M, Brian D Ziebart, James Andrew Bagnell, and Martial Hebert (2012). “Activity forecasting”. In: *European Conference on Computer Vision*. Springer, pp. 201–214 (cit. on pp. 7, 8, 58, 75, 76, 78, 79).

Köhler, Sebastian, Michael Goldhammer, Sebastian Bauer, et al. (2012). “Early detection of the pedestrian’s intention to cross the street”. In: *2012 15th International IEEE Conference on Intelligent Transportation Systems*. IEEE, pp. 1759–1764 (cit. on p. 22).

Kohonen, Teuvo (Jan. 1982). “Self-organized formation of topologically correct feature maps”. In: *Biological Cybernetics* 43.1, pp. 59–69 (cit. on p. 89).

Kooij, Julian FP, Nicolas Schneider, and Dariu M Gavrila (2014). “Analysis of pedestrian dynamics from a vehicle perspective”. In: *2014 IEEE Intelligent Vehicles Symposium Proceedings*. IEEE, pp. 1445–1450 (cit. on p. 19).

Lankton, Peter (2007). *Endlsey’s model of SA* (cit. on p. 4).

Lefèvre, Stéphanie, Dizan Vasquez, and Christian Laugier (July 2014). “A survey on motion prediction and risk assessment for intelligent vehicles”. In: *ROBOMECH Journal* 1.1, p. 1 (cit. on p. 17).

Liu, Ce, Jenny Yuen, and Antonio Torralba (2009). “Nonparametric scene parsing: Label transfer via dense scene alignment”. In: Institute of Electrical and Electronics Engineers (cit. on p. 8).

Liu, N., R. I. A. Davis, B. C. Lovell, and P. J. Kootsookos (Apr. 2004). “Effect of initial HMM choices in multiple sequence training for gesture recognition”. In: *International Conference on Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004.* Vol. 1, 608–613 Vol.1 (cit. on p. 51).

Luber, Matthias, Gian Diego Tipaldi, and Kai O Arras (2011). “Place-dependent people tracking”. In: *The International Journal of Robotics Research* 30.3, pp. 280–293. eprint: https://doi.org/10.1177/0278364910393538 (cit. on p. 82).

Lv, Fengjun and Ramakant Nevatia (2006). “Recognition and segmentation of 3-d human action using hmm and multi-class adaboost”. In: *European conference on computer vision*. Springer, pp. 359–372 (cit. on p. 34).

Madigan, David, Jeremy York, and Denis Allard (1995). “Bayesian Graphical Models for Discrete Data”. In: *International Statistical Review / Revue Internationale de Statistique* 63.2, pp. 215–232 (cit. on p. 30).

Martinetz, Thomas and Klaus Schulten (1991). "A" neural gas" network learns topologies, Artificial Neural Networks". In: *Proceedings of the 1991 International Conference, ICANN-91*. Vol. 1, pp. 397–402 (cit. on p. 44).

McFarland, David (1989). *Problems of animal behaviour*. Longman Sc & Tech (cit. on p. 18).

McKnight, A James and Bert B Adams (1970a). "Driver education task analysis. Volume 1: Task descriptions". In: (cit. on p. 4).

– (1970b). "Driver Education Task Analysis. Volume II: Task Analysis Methods. Final Report." In: (cit. on p. 4).

McKnight, A James and Alan G Hundt (1971). "Driver education task analysis: volume III: Instructional objectives". In: (cit. on p. 4).

Montel, Marie Claude, Thierry Brenac, Marie-Axelle Granie, Marine Millot, and Cécile Coquelet (2013). "Urban environments, pedestrian-friendliness and crossing decisions". In: *Transportation Research Board 92nd Annual Meeting*, 13p (cit. on pp. 50, 63, 66, 68).

Moravec, H. and A. Elfes (Mar. 1985). "High resolution maps from wide angle sonar". In: *Proceedings. 1985 IEEE International Conference on Robotics and Automation*. Vol. 2, pp. 116–121 (cit. on p. 56).

Munoz, Daniel, J. Andrew Bagnell, and Martial Hebert (2010). "Stacked Hierarchical Labeling". In: *Computer Vision – ECCV 2010*. Ed. by Kostas Daniilidis, Petros Maragos, and Nikos Paragios. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 57–70 (cit. on p. 108).

National Transportation Safety Board (May 2018). *PRELIMINARY REPORT HIGHWAY HWY18MH010*. Tech. rep. (cit. on p. 3).

Neal, Radford M and Geoffrey E Hinton (1998). "A view of the EM algorithm that justifies incremental, sparse, and other variants". In: *Learning in graphical models*. Springer, pp. 355–368 (cit. on pp. 48, 51).

Oh, S., A. Hoogs, A. Perera, et al. (June 2011). "A large-scale benchmark dataset for event recognition in surveillance video". In: *CVPR 2011*, pp. 3153–3160 (cit. on p. 75).

P. Blom, H. A. (Dec. 1984). "An efficient filter for abruptly changing systems". In: *The 23rd IEEE Conference on Decision and Control*, pp. 656–658 (cit. on p. 18).

Pacione, Michael (1980). "REDEVELOPMENT OF A MEDIUM-SIZED CENTRAL SHOPPING AREA A CASE STUDY OF CLYDEBANK". In: *Tijdschrift voor economische en sociale geografie* 71.3, pp. 159–168 (cit. on p. 62).

Pellegrini, Stefano, Andreas Ess, Konrad Schindler, and Luc Van Gool (2009). "You'll never walk alone: Modeling social behavior for multi-target tracking". In: *2009 IEEE 12th International Conference on Computer Vision*. IEEE, pp. 261–268 (cit. on p. 19).

Peponis, John, Craig Zimring, and Yoon Kyung Choi (1990). "Finding the building in wayfinding". In: *Environment and behavior* 22.5, pp. 555–590 (cit. on p. 62).

Pérez-Hurtado, Ignacio, Jesús Capitán, Fernando Caballero, and Luis Merino (2015). "An extension of GHMMs for environments with occlusions and automatic goal discovery for person trajectory prediction". In: *European Conference on Mobile Robots (ECMR)*. IEEE, pp. 1–7 (cit. on pp. 82, 83, 85, 89, 91, 94, 97, 109, 112, 117, 125, 128, 129, 138).

Qi, Charles R, Hao Su, Kaichun Mo, and Leonidas J Guibas (2017). "Pointnet: Deep learning on point sets for 3d classification and segmentation". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 652–660 (cit. on p. 123).

Quintero, Raúl, Ignacio Parra, J Lorenzo, D Fernández-Llorca, and MA Sotelo (2017). "Pedestrian intention recognition by means of a hidden markov model and body language". In: *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, pp. 1–7 (cit. on p. 23).

Rabiner, Lawrence R, B-H Juang, SE Levinson, and MM Sondhi (1985). "Some properties of continuous hidden Markov model representations". In: *AT&T technical journal* 64.6, pp. 1251–1270 (cit. on p. 50).

Rabiner, Lawrence R and Biing-Hwang Juang (1993). *Fundamentals of speech recognition*. Vol. 14. PTR Prentice Hall Englewood Cliffs (cit. on pp. 34, 36, 50).

Rasouli, A., I. Kotseruba, and J. K. Tsotsos (June 2017). "Agreeing to cross: How drivers and pedestrians communicate". In: *2017 IEEE Intelligent Vehicles Symposium (IV)*, pp. 264–269 (cit. on p. 142).

Redmon, Joseph and Ali Farhadi (2016). "YOLO9000: Better, Faster, Stronger". In: *arXiv preprint arXiv:1612.08242* (cit. on p. 107).

– (2018). "YOLOv3: An Incremental Improvement". In: *CoRR* abs/1804.02767. arXiv: 1804.02767 (cit. on p. 5).

Rehder, Eike and Horst Kloeden (2015). "Goal-directed pedestrian prediction". In: *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 50–58 (cit. on p. 25).

Reif, John, John Reif, and Micha Sharir (July 1994). "Motion Planning in the Presence of Moving Obstacles". In: *J. ACM* 41.4, pp. 764–790 (cit. on p. 1).

Ridel, D., E. Rehder, M. Lauer, C. Stiller, and D. Wolf (Nov. 2018). "A Literature Review on the Prediction of Pedestrian Behavior in Urban Scenarios". In: *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 3105–3112 (cit. on pp. 8, 17, 106).

Robles, Bernard, Manuel Avila, Florent Duculty, et al. (2012). "Methods to choose the best Hidden Markov Model topology for improving maintenance policy". In: *9th International Conference on Modeling, Optimization & SIMulation* (cit. on p. 36).

Rothenbücher, D., J. Li, D. Sirkin, B. Mok, and W. Ju (Aug. 2016). "Ghost driver: A field study investigating the interaction between pedestrians and driverless vehicles". In: *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 795–802 (cit. on p. 102).

Rudenko, Andrey, Luigi Palmieri, Michael Herman, et al. (2019). "Human Motion Trajectory Prediction: A Survey". In: *arXiv preprint arXiv:1905.06113* (cit. on p. 17).

Rummelhard, L., A. Nègre, and C. Laugier (Sept. 2015). "Conditional Monte Carlo Dense Occupancy Tracker". In: *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, pp. 2485–2490 (cit. on pp. 56, 124).

SAE International (2018). *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles* (cit. on pp. 2, 3).

Schneider, Nicolas and Dariu M. Gavrila (2013). "Pedestrian Path Prediction with Recursive Bayesian Filters: A Comparative Study". In: *Pattern Recognition*. Ed. by Joachim Weickert, Matthias Hein, and Bernt Schiele. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 174–183 (cit. on pp. 55, 106).

Schulz, Andreas T and Rainer Stiefelhagen (2015). "A controlled interactive multiple model filter for combined pedestrian intention recognition and path prediction". In: *2015 IEEE 18th International Conference on Intelligent Transportation Systems*. IEEE, pp. 173–178 (cit. on p. 23).

Schulz, Andreas Th and Rainer Stiefelhagen (2015). "Pedestrian intention recognition using latent-dynamic conditional random fields". In: *2015 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, pp. 622–627 (cit. on p. 22).

Soruri, M., S. Hamid Zahiri, and J. Sadri (Mar. 2013). "A new approach of training Hidden Markov Model by PSO algorithm for gene Sequence Modeling". In: *2013 First Iranian Conference on Pattern Recognition and Image Analysis (PRIA)*, pp. 1–4 (cit. on p. 36).

Su, Hang, Jun Zhu, Yinpeng Dong, and Bo Zhang (2017). "Forecast the Plausible Paths in Crowd Scenes." In: *IJCAI*. Vol. 1, p. 2 (cit. on p. 25).

Svenstrup, M., T. Bak, and H. J. Andersen (Oct. 2010). "Trajectory planning for robots in dynamic human environments". In: *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 4293–4298 (cit. on pp. 56, 64).

Tesla, Inc (2019). *Tesla to host Autonomy Day*. Ed. by Tesla Motors (cit. on p. 2).

Thompson, Leah L, Frederick P Rivara, Rajiv C Ayyagari, and Beth E Ebel (2013). "Impact of social and technological distraction on pedestrian crossing behaviour: an observational study". In: *Injury prevention* 19.4, pp. 232–237 (cit. on p. 93).

Thrun, Sebastian, Wolfram Burgard, and Dieter Fox (2005). *Probabilistic robotics* (cit. on pp. 27, 42, 56).

Turner, Alasdair and Alan Penn (2002). "Encoding Natural Movement as an Agent-Based System: An Investigation into Human Pedestrian Behaviour in the Built Environment". In: *Environment and Planning B: Planning and Design* 29.4, pp. 473–490. eprint: https://doi.org/10.1068/b12850 (cit. on p. 81).

Varadarajan, Jagannadan and Jean-Marc Odobez (2009). "Topic models for scene analysis and abnormality detection". In: *IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops)*. IEEE, pp. 1338–1345 (cit. on pp. 20, 106, 107).

Vasishta, Pavan, Dominique Vaufreydaz, and Anne Spalanzani (2017). "Natural Vision Based Method for Predicting Pedestrian Behaviour in Urban Environments". In: *IEEE 20th International Conference on Intelligent Transportation Systems* (cit. on p. 81).

Vasko, Raymond C, Amro El-Jaroudi, and J Robert Boston (1996). "An algorithm to determine hidden Markov model topology". In: *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*. Vol. 6. IEEE, pp. 3577–3580 (cit. on p. 36).

Vasquez, D. (2007). "Incremental learning for motion prediction of pedestrians and vehicles". PhD thesis. Grenoble INPG (cit. on pp. 17, 27, 45, 46).

– (May 2016). "Novel planning-based algorithms for human motion prediction". In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3317–3322 (cit. on pp. 24, 58).

Vasquez, D., T. Fraichard, and C. Laugier (2009). "Incremental Learning of Statistical Motion Patterns With Growing Hidden Markov Models". In: *IEEE Transactions on Intelligent Transportation Systems* 10.3, pp. 403–416 (cit. on pp. 7, 10, 29, 39, 41–43, 50, 89, 91, 97).

Viterbi, A. (Apr. 1967). "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm". In: *IEEE Transactions on Information Theory* 13.2, pp. 260–269 (cit. on p. 34).

Volpe, Richard and Pradeep Khosla (1993). "A theoretical and experimental investigation of impact control for manipulators". In: *The International Journal of Robotics Research* 12.4, pp. 351–365 (cit. on p. 68).

Völz, Benjamin, Karsten Behrendt, Holger Mielenz, et al. (2016). "A data-driven approach for pedestrian intention estimation". In: *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, pp. 2607–2612 (cit. on p. 142).

Vondrick, Carl, Donald Patterson, and Deva Ramanan (2012). "Efficiently Scaling up Crowdsourced Video Annotation". In: *International Journal of Computer Vision*. 10.1007/s11263-012-0564-1, pp. 1–21 (cit. on pp. 108, 123).

Wada, Kentaro (2018). *labelme: Image Polygonal Annotation with Python*. https://github.com/wkentaro/labelme (cit. on p. 75).

Wakabayashi, Daisuke (2018). "Self-Driving Uber Car Kills Pedestrian in Arizona, Where Robots Roam". In: *The New York Times*, A1 (cit. on p. 3).

Wang, Shu, Huchuan Lu, Fan Yang, and Ming-Hsuan Yang (2011). "Superpixel tracking". In: *2011 International Conference on Computer Vision*. IEEE, pp. 1323–1330 (cit. on p. 75).

Wang, Zhan, Rares Ambrus, Patric Jensfelt, and John Folkesson (2014). "Modeling motion patterns of dynamic objectsby IOHMM". In: *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE conference proceedings, pp. 1832–1838 (cit. on p. 82).

Wang, Zhan, Patric Jensfelt, and John Folkesson (2016). "Building a human behavior map from local observations". In: *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, pp. 64–70 (cit. on p. 58).

Weiss, T., B. Schiele, and K. Dietmayer (June 2007). "Robust Driving Path Detection in Urban and Highway Scenarios Using a Laser Scanner and Online Occupancy Grids". In: *2007 IEEE Intelligent Vehicles Symposium*, pp. 184–189 (cit. on p. 56).

Wojke, Nicolai, Alex Bewley, and Dietrich Paulus (2017). "Simple Online and Realtime Tracking with a Deep Association Metric". In: *arXiv preprint arXiv:1703.07402* (cit. on p. 107).

Wolf, M. T. and J. W. Burdick (May 2008). "Artificial potential functions for highway driving with collision avoidance". In: *2008 IEEE International Conference on Robotics and Automation*, pp. 3731–3736 (cit. on pp. 56, 64, 65, 70, 71).

World Health Organisation (Dec. 2018). *Global status report on road safety 2018*. Research rep. (cit. on p. 1).

Xu, M. and Y. Wang (Jan. 2016). "Quantifying PM2.5 Concentrations From Multi-Weather Sensors Using Hidden Markov Models". In: *IEEE Sensors Journal* 16.1, pp. 22–23 (cit. on p. 36).

Yamaguchi, Kota, Alexander C Berg, Luis E Ortiz, and Tamara L Berg (2011). "Who are you with and where are you going?" In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, pp. 1345–1352 (cit. on p. 19).

Zhang, Feihu, Daniel Clarke, and Alois Knoll (2014). "Vehicle detection based on lidar and camera fusion". In: *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. IEEE, pp. 1620–1625 (cit. on p. 5).

Zhou, Yin and Oncel Tuzel (2017). "VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection". In: *CoRR* abs/1711.06396. arXiv: 1711.06396 (cit. on p. 123).

Ziebart, Brian D (2010). "Modeling purposeful adaptive behavior with the principle of maximum causal entropy". In: (cit. on p. 78).

Ziebart, Brian D, Andrew L Maas, J Andrew Bagnell, and Anind K Dey (2008). "Maximum Entropy Inverse Reinforcement Learning." In: *AAAI*. Vol. 8. Chicago, IL, USA, pp. 1433–1438 (cit. on pp. 10, 50, 57).

Ziebart, Brian D, Nathan Ratliff, Garratt Gallagher, et al. (2009). "Planning-based prediction for pedestrians". In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 3931–3936 (cit. on pp. 7, 57, 58, 78).