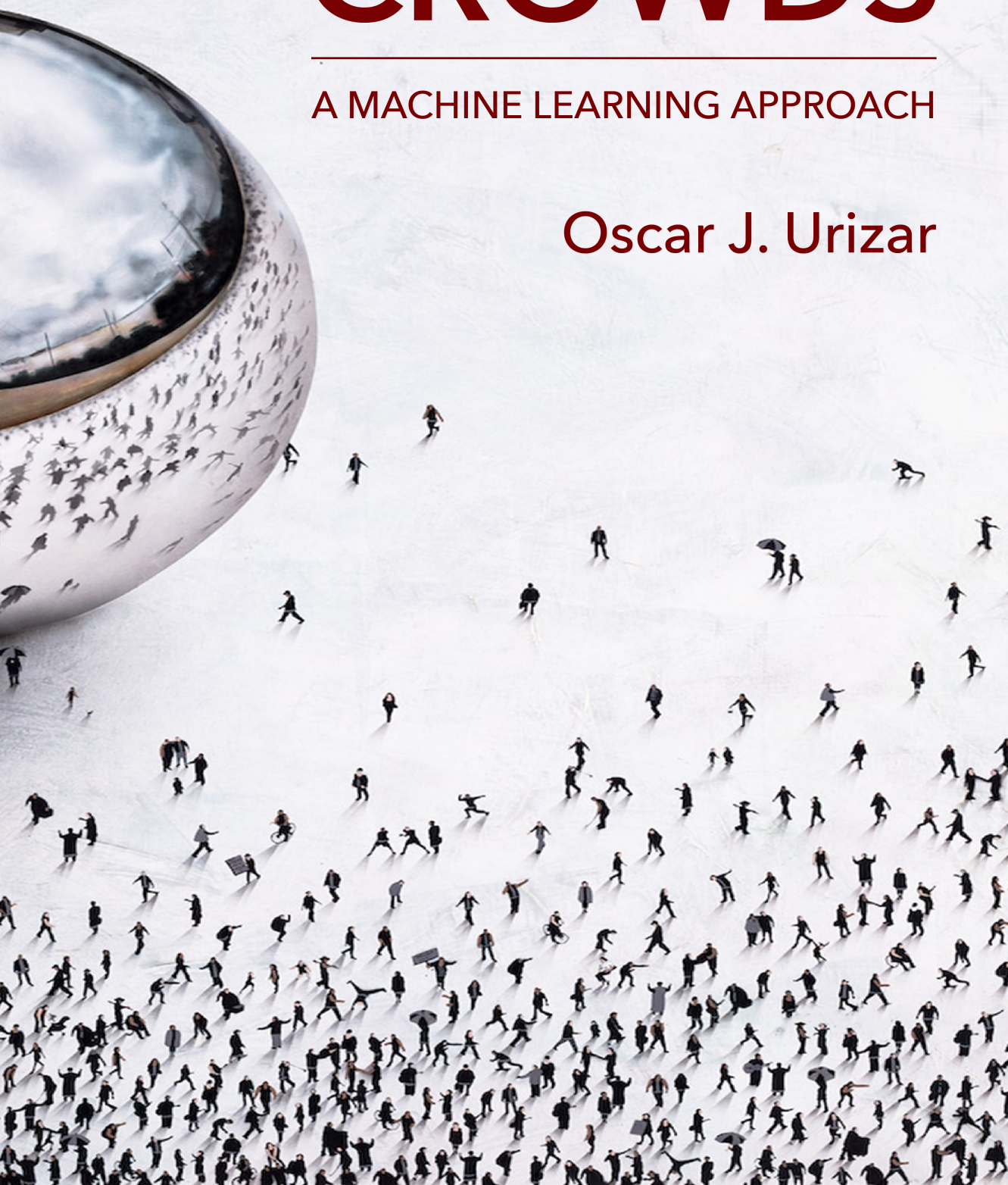
Emotion estimation in crowds

EMOTION ESTIMATION IN CROWDS

A MACHINE LEARNING APPROACH

Oscar J. Urizar



Emotion Estimation in Crowds: A Machine Learning Approach

Oscar J. Urizar

CIP-DATA LIBRARY TECHNISCHE UNIVERSITEIT EINDHOVEN

JUAREZ URIZAR, O.R.

Emotion Estimation in Crowds: A Machine Learning Approach by Oscar J. Urizar.
Eindhoven: Technische Universiteit Eindhoven, 2020. Proefschrift.

A catalogue record is available from the Eindhoven University of Technology Library

ISBN 978-90-386-5099-9

Keywords: affective models, emotion estimation, crowd emotions, crowd analysis

Emotion Estimation in Crowds: A Machine Learning Approach

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de rector magnificus prof.dr.ir. F.P.T. Baaijens, voor een commissie aangewezen door het College voor Promoties, in het openbaar te verdedigen op donderdag 29 oktober 2020 om 16:00 uur

door

OSCAR RICARDO JUAREZ URIZAR

geboren te Guatemala, Guatemala

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

voorzitter:	prof.dr. L. Chen
1e promotor:	prof.dr. C.S. Regazzoni (Università degli studi di Genova)
2e promotor:	prof.dr. G.W.M. Rauterberg
co-promotor:	dr. L. Marcenaro (Università degli studi di Genova)
co-promotor:	dr.ir. E.I. Barakova
externe leden:	prof.dr. F. Toschi
	prof.dr. S.B. Vos
	prof.dr. T. Bosse (RUN)

UNIVERSITÀ DEGLI STUDI
DI GENOVA

TU/e Technische Universiteit
Eindhoven
University of Technology

This dissertation was produced under Erasmus Mundus Joint Doctorate Program in Interactive and Cognitive Environments. The research was conducted towards a joint double PhD degree between the following partner universities:

Università degli Studi di Genova

&

Technische Universiteit Eindhoven

Acknowledgments

This PhD Thesis has been developed in the framework of, and according to, the rules of the Erasmus Mundus Joint Doctorate on Interactive and Cognitive Environments EMJD ICE [FPA n° 2010-0012] with the cooperation of the following Universities:

Alpen-Adria-Universität Klagenfurt – AAU

Queen Mary, University of London – QMUL

Technische Universiteit Eindhoven – TU/e

Università degli Studi di Genova – UNIGE

Universitat Politècnica de Catalunya – UPC

According to ICE regulations, the Italian PhD title has also been awarded by the Università degli Studi di Genova.

To Laura, Mercedes and Vicente.

Gratitude

The country where I was born, Guatemala, faces immeasurable obstacles to provide even the most basic education to its children, and so, access to higher education is a luxury out of reach for more than ninety percent of its population. Venturing abroad in the pursue of a doctorate degree, as you can imagine now, is close to impossible. As I learned over the years, a feat of this magnitude requires an outstanding intellect and determination, or as it was my case, hard work, perseverance, and a peculiar group of people willing to give you the support and opportunity to succeed. I would like to express my sincere gratitude to all those who made it possible for me to realize this dream.

To my family
Thank you for your overwhelming and unconditional support from the very moment I decided to start this beautiful journey. Your example of perseverance, kindness and humility lead me to where I find myself now.

To my professors
I would like to express my sincere gratitude for giving me the opportunity to join this program. To Prof. Carlo S. Regazzoni, grazie mile for your guidance, patience, and good sense of humor in all our conversations. To Prof. Matthias Rauterberg, thank you for your advice and honest feedback, it was very helpful in guiding my research. To Prof. Emilia Barakova, I sincerely appreciate all your active support and advice; you were always generous with your time, helping me to move forward. To Prof. Lucio Marcenaro, thank you for always been ready to help in technical and practical matters.

To my colleagues
During these years of PhD life, I was lucky enough to be surrounded by helpful and kind colleagues with whom we shared our frustrations and joys. Thanks to Baydoun and Mahdyar for all the fun times we had in Genova. A big thank you to Damian Campo for becoming such a good friend, helping me in my research by day, and entertaining so many interesting and pointless conversations by night.

To my friends
I am happy to say I found much support from old and new friends in the process of completing this doctorate. Thank you to Damian Głowienka, the friend I found by accident and whom was always there to share a beer and talk out our daily struggles. Thank you to Cynthia Lin, my true and unconditional friend, always looking after me regardless of the distance. Finally, I would like to thank Sophie Blommerde for giving me that last push at the end of my PhD life, your unconditional support and care helped me more than you know.

Oscar J. Urizar

Netherlands, April 2020.

Tutto quello che vuoi, e fu quello il saluto.
Tutto quello che voglio, alla fine, l'ho avuto.

Abstract

Before 1830, no one really felt emotions. Instead, they felt 'passions', 'accidents of the soul', 'moral sentiments', and explained them very differently from how we conceive emotions today. We then began to unveil the inner workings of human affects and their strong impact on behavior, evidencing the fundamental role of emotions in understanding crowds. With cities and urban areas growing at an increasing pace, the development of intelligent systems capable to identify emotions in crowds will prove to be of great value in ensuring safety, stability and efficient management of crowds in public areas.

Incorporating our current knowledge in the fields of machine learning, crowd behavior analysis, and psychology, this research presents a method to infer the emotional states of individual pedestrians and the crowd as a whole by analyzing walking trajectories captured via surveillance cameras. This is accomplished by first building data-driven behavior models capable of describing the dynamics of both pedestrians and the crowd, and second, by learning the contextual association between the observed behavior and the underlying emotion. The behavior models are constructed using dynamic Bayesian networks that capture the influence between the detected movement and the motivation driving such actions. The emotional state of either the pedestrian or the crowd is inferred by identifying its motivation and quantifying the deviation between observed and expected behavior.

A series of experiments are conducted at different stages of development for the pedestrian and crowd model to assess their validity and accuracy. Chapter 5 presents the first iteration of the pedestrian model and an experiment is conducted employing a simulation tool to produce a virtual crowded environ-

ment where pedestrians interact in a complex environment to reach a particular destination. In this first experiment, our model is evaluated for its predictive capability to infer the desired destination (i.e., motivation) and short term movement of individual pedestrians. In chapter 6, the elements of expectation and emotion are incorporated; and an experiment combining both real-world and simulated datasets is conducted to confirm the model's capability to infer pedestrian emotions. Our first version of the crowd model is introduced in chapter 7 where an experiment using a real-world dataset evaluates the model's ability to describe the behavior of multiple pedestrians as a single entity. Finally in chapter 8, an improved version of the crowd model is provided, accompanied by an experiment that tests our model in a wide variety of scenarios.

The models developed in this thesis provide several contributions. The pedestrian model presents a data-driven model based on hierarchical Bayesian networks that includes multiple levels of abstraction to account for behavioral and psychological factors, and is capable to generalize to different contexts in a supervised way; it introduces a distance-to-motivation (distance-to-motivation (DTM)) measurement which helps to form an association between observable behavior and emotions with strong foundation in psychological principles; and finally, we develop an emotion annotation scheme for automatic labeling of pedestrian trajectories based on learned motivations and expectations. Similarly, the crowd model implements a data-driven model based on hierarchical Bayesian networks capable to generalize to ambulatory crowds in multiple contexts, and that expands the concepts of motivation, expectation and emotions from the individual pedestrian to a collective level in a consistent and equivalent way for both the pedestrian and the crowd; it incorporates a method to describe crowds and sub-crowds based on spatial-temporal interactions learned from partial observation of pedestrian trajectories; and it accounts collective emotion of crowds and sub-crowds measured in a continuum valence axis, in a way that is representative and consistent with the emotions experienced by the pedestrians member of the crowd.

Future work in the direction of emotion estimation in crowds can be focused in addressing limitations encountered throughout our research. The first and foremost limitation is that the models presented in this thesis apply only to crowds with a predominantly ambulatory behavior, as is the case of casual crowds, queues, acquisitive, mobs, riots, and panic crowds. Since we employ surveillance cameras to observe the crowd due to their ubiquitousness in public spaces, we depend on the performance of crowd counting and pedestrian detection & tracking algorithms. This entails that in the presence of highly dense crowds, pedestrian detection & tracking algorithms will deliver fragmented tra-

jectories, tampering the performance of the pedestrian model. Another relevant limitation concerns the absence of context-awareness in our method, i.e., the association of learned behavior to an emotion label is dependent on the context of the situation. For this reason, models of behaviors learned unsupervisedly need to be empirically labeled by a human operator.

Contents

List of Figures

List of Tables

Part I

PREAMBLE

Chapter 1

Introduction

1.1 Motivation

On July 14, 1789, after enduring long years of calamity and heavy taxation schemes, a crowd of Parisians stormed the Bastille, marched towards Versailles, and overthrew king Louis XVI, initiating the French Revolution [12]. In the following years, crowds where at the center of numerous violent events, yet these crowds were not composed by soldiers or militia, nor criminals, but by everyday people who ignited by the amplified emotions in the collective, willingly engaged in acts of brutality. Later studies of these and similar events started to point to the notion that people feel and behave differently when they become part of a crowd. Gustav Le Bon shares this thought in his famous book *The Crowd, A Study of the Popular Mind* [74]: "The most striking peculiarity presented by a crowd is the following: Whoever be the individuals that compose it, however like or unlike be their mode of life, their occupations, their character, or their intelligence, the fact that they have been transformed into a crowd puts them in possession of a sort of collective mind which makes them feel, think, and act in a manner quite different from that in which each individual of them would feel, think, and act were he is in a state of isolation". It was also Le Bon among the first scholars to point out that crowds can be prompted into actions of heroism and goodness as much as to destruction and violence, the main driver in behavior is their underlying emotions. This assertion that emotions play an essential role in regulating the behavior of crowds is strongly supported by a vast amount of research in the field of social psychology, with prominent

exponents in the likes of Le Bon [74], McDougall [66], and Freud [54].

The advancement in our knowledge of emotions and human affects had a journey similar to that of crowds, coming from obscure hypothesis to formally defined and widely accepted theories. T. Smith eloquently says, "No one really felt emotions before about 1830. Instead, they felt other things-'passions', 'accidents of the soul', 'moral sentiments'-and explained them very differently from how we understand emotions today" [122]. In the contemporary literature regarding emotions, three leading contenders prevail among the many proposed theories: (a) Discrete emotion theories, (b) dimensional theories, and (c) appraisal theories [114]. Over the past decades, we have witnessed a significant departure from theoretical to empirical research, helping us to paint a more clear picture of the inner workings of emotions, up to the point where we can dissect, measure, and test for emotions. Some examples in this direction are Paul Ekman's facial action coding system [46] aiming to explain the universality of emotions expressed in facial expressions, Antonio Damasio's somatic marker hypothesis [60] attempting to understand how emotions and their underlying biology influence decision making, and many others.

Having this knowledge of emotions and their influence in crowds at our disposal started to prove beneficial as we were able to plan better strategies to manage crowds at big gatherings [13] [140] and prevent fatal incidents as it has been the case with numerous examples in the past [42]. Another critical leap forward in producing practical applications came with the birth of Affective Computing, outlined by Rosalind Picard's 1995 paper on this topic. A relatively new branch of computer science that aims to breach the boundaries among the fields of psychology, neuroscience, and computer science. In its purest form, affective computing employs our current understanding of emotions, not limited to crowds, and attempts to build models capable of identifying, process, and mimic human affect. Useful examples under this branch of science allowed us to automate tasks of recognizing emotions from facial expressions, body language, and even by the voice of people [132]. Under the scope of crowds, the area of Crowd Analysis is already mature enough to accomplish the tasks of counting and tracking pedestrians, modeling crowd dynamics, and identifying abnormalities, up to far more complex concepts like recognizing interactions and predicting their behavior [34] [141]. Looking into the future, smart cities, ambient intelligence, and advanced cognitive dynamic systems are starting to go beyond safety concerns to the enhancement of people's experience in smart environments [33] [9].

At the intersection of affective computing and crowd analysis, with a foundation in psychology, this research aims to address the gap in the literature

regarding appropriate methods to infer emotional states of individual pedestrians and the crowd as a whole by analyzing walking trajectories captured via surveillance cameras. This is accomplished by first building data-driven behavior models capable of describing the dynamics of both pedestrians and the crowd, and second, by learning the contextual association between the observed behavior and the underlying emotion. The behavior models are constructed using dynamic Bayesian networks that capture the influence between the detected movement and the motivation driving such actions. The emotional state of either the pedestrian or the crowd is inferred by identifying its motivation and quantifying the deviation between observed and expected behavior. Having methodologies of this nature at our disposal has and will be valuable as we move towards bigger and more crowded cities, enabling us to better design spaces and systems capable of procuring a positive and safe experience in crowded environments.

1.2 Theoretical Framework

The method proposed in this research operates under the premise that pedestrians travel across an environment motivated by the desire to reach a destination, and the cognitive process determining how to navigate towards their intended motivation is influenced by their emotional state, interaction with other people, and situational constraints. In this manner, the ambulatory actions exhibited by a person can be considered as emotional responses to (internal or external) stimuli, comparable to the way physiological changes like sudden cheek blush, variations in perspiration, or shifts in heart rate are acknowledged to be linked to emotional states.

The first point of reference for the work presented in this thesis comes from Behaviorism, proposing the idea of interpreting observable behavior as a relation between stimuli and responses. More specifically, Behaviorism is the school of thought in psychology that pertained to see human behavior as automatic responses to stimuli over which emotions had an important role. Near the year of 1863, Ivan Sechenov steered this line of inquiry under the presumption that these "automatic" behavioral responses were part of a protective mechanism for our body [22]. Years later, Ivan Pavlov would add significant contributions furthering our knowledge in classical conditioning, a learning process by which a biologically potent stimulus is paired to a neutral stimulus up until the point where the neutral stimulus alone is capable of eliciting a behavioral response quite similar to the one obtained by the potent stimulus. Many more significant

contributions would follow, helping to expand the foundation of this theory. The advantage of Behaviorism resided in its focus on observable behavior, as it is easy to quantify and collect data from, allowing to provide explanations and clear evidence for the experiments conducted. On the downside, Behaviorism is criticized for its superficiality as it only focuses on what is measurable by observation, failing to account for the complex and still elusive internal processes in our brain and body.

The second reference to support this view is borrowed from Field Theory, developed by Kurt Lewis and Gestald [20]. This theory aims to explain the mechanisms of interaction between individuals and the field they currently occupy (i.e., their environment). In this context, the individual is conceptualized to have *motivations* driven by personal *needs*, *beliefs* concerning its own state as well as the state of the environment, and *abilities* to interact with the environment and other individuals. The environment is the physical space and the context a person perceives and acts on. Behavior encompasses the set of actions a person performs, which causes a change in the environment and other people, contemplating that perceived changes in the environment will, in turn, influence the subsequent behavior of that same person, in what is known as the perception-action cycle [32]. A consequential element to behavior and motivation is *expectation*, which provides an assessment of the effort required to meet a person's motivations. The framing of this theory is suitable for our study as it provides the building blocks of behaviors, motivations, and expectations.

The final point of support concerns the study in the cognitive process of Decision Making [97]. In its purest form, decision making is a continuous cognitive process that enables a person to interact with the environment utilizing perceiving changes in the environment and responding accordingly by identifying and selecting beliefs, decisions, and potentially take actions. Differing from Behaviorism, decision making accounts for internal factors like biases, personality traits, constrains by social normative, and emotions. Presented with constant uncertainty about the choices that are beneficial or harmful to us, emotions appear to take a central role in facilitating the decision-making process that will ensure our survival. In the particular case of crowded environments, the complex and apparently chaotic interaction among a high number of people prevents single individuals from perceiving enough information about their surroundings. Such conditions prompt our brain to make use of Affect Heuristics, a subconscious process, a mental shortcut that allows us to make decisions quickly with limited understanding of the context. Emotions remain a constant guardian of our behavioral responses, from the subtle and mundane moments of our daily life to the more extreme situations like panic crowds running away

from a violent event or spectators crowds bursting into excitement at a sport event.

The composition of our theoretical foundation argues that emotions play a central role in guiding a person's behavior in a way that is consistent to personal beliefs and motivations. However, the above presented references draw knowledge from apparently conflicting stands. The first prospect attempts to explain emotions as a mechanistic product between received stimuli and the expressed behavioral responses, pointing towards a *radical behaviorism* approach. The second and third points of reference address the importance of personal beliefs and motivations in the elicitation of emotions, leaving us with a perspective captured by *intentional psychology*. Previous contributions have proposed resolutions to this conflict, namely, Foxall [53] and Goleman et al. [84] appeal to the complementary role between these two notions as they both, to a certain extent, have explanatory merits. The consensus stems from the acknowledgement that whereas behavior is an important facet of emotions, its interpretation can differ depending on the circumstances around each individual, hence we also require an understanding of the context presented to the individual before we can confidently proceed to identify emotions. This allusion to contextual relation between behavior and emotion is a concern accounted for throughout this thesis where our proposed methodology learns behavioral models in an unsupervised way but the association to emotions requires supervised intervention. Consequently, an inherent limitation is that models are built and applicable only for a specific environment and context. Another relevant limitation in our approach, lies in the fact that we account only for emotions manifested in behaviors where the corresponding motivations relate to reaching a destination in the environment. Conversely, the emotions elicited by factors unrelated to the individual's presence and interaction with the environment are not identifiable by our approach. Despite this theoretical argument and the identified limitations, the fundamental issue is then whether embodied behaviors provide sufficient enough cues to infer emotions, and to this point existing literature provides numerous precedents of in our favor. For instance, the work presented in [81] explores how can emotions be grounded in the embodiment of robots to facilitate interaction with humans. Another example is found in [100] where the sentiment and engagement of patients with moderate and advanced dementia is inferred by Laban analysis of movement as sufferers of this condition can not self-reflect, and some rarely speak.

After careful consideration of the previously presented and other relevant literature to be introduced in the following chapters, we believe to have a solid theoretical foundation to propose a methodology for inferring emotions

of pedestrians at the individual and collective levels from observable behavior. However, as new research expands the body of knowledge, we continue to confirm the complexity of the human psyche; therefore it is vital to recognize the limitations in using the literature of Field Theory, Behaviorism and Decision Making as the basis for our study of crowds. With this consideration in mind, our approach in this thesis aims to provide a restricted but reasonable approximation of the emotional state of pedestrians and the crowd given the limited information we can observe and measure.

1.3 Crowd

Due to the central role that crowds take in the research presented here, this section aims to clearly establish the characteristics and concise definition to be used in this work.

1.3.1 Previous Definitions

The term *crowd* is loosely used to denote a large group of people, taking for granted what does and does not constitute a crowd. This is the case even in academic literature, where scholars appear to disagree on a single concise definition. To illustrate this lack of consensus, we list several definitions previously used:

- Le Bon, 1897 [74]: *“an agglomeration of men presents new characteristics very different from those of the individuals composing it. The sentiments and ideas of all the persons in the gathering take one and the same direction, and their conscious personality vanishes.”*
- Tilly, 1978 [126]: "*people acting together in the pursuit of common interests.*"
- Lofland, 1985 [80]: "*a large number of people in the same place at the same time.”*
- McPhail, 1991 [87]: "*two or more persons engaged in one or more behaviors judged common or concerted in one or more dimensions.”*
- Musse & Thalmann, 1997 [94]: "*a large group of individuals in the same physical environment, sharing a common goal (e.g., people going to a rock*

show or a football match). The individuals in a crowd may act in a different way than when they are alone or in a small group."

- Brown & Lewis, 1998 [19]: *"a compact gathering or collection of people with connotations of homogeneity of characteristics and unanimity of behavior."*
- Sharma, 2000 [118]: *"they are present in a common environment, and all the individuals present in the crowd usually share a common goal."*
- Myers, 2005 [95]: *"two or more people who, for longer than a few moments, interact with and influence one another and perceive one another as 'us'."*
- Willems et al. [2]: *"A (typically large) number of people in one place at the same time. It is possible that a physical crowd contains one or more psychological crowds (e.g. football fans in a transport hub with commuters).."*

1.3.2 Features

On a closer examination of the proposed definitions, it soon comes to the realization of a series of common features for defining what we mean when we talk about crowds [24]. As follow, we list these key features in alphabetical order along with their corresponding definition.

- **Collectivity**: It encompasses social identity, goals, interests, and behaviors. It is this feature that sparks interest in studying collective behavior, collective emotions, and other aspects in the field of Social Psychology. More importantly, this criteria draws a distinction between physical crowds and psychological crowds [44, 107]. A Physical crowd, sometimes referred to as a mass, expresses no collectivity among its members. Oppositely, collectiveness is a fundamental part of psychological crowds.
- **Density**: The size of a crowd becomes relevant only under a certain density level. A hundred people spread across a large park (low density) would not constitute a crowd, but the same number assembled in few meters of that same park (high density) would. As outlined in [63, 98], we identify three density levels under which an aggregate of people can rightfully be considered a crowd. Each of these density levels is illustrated in Figure 1.1.
- **Interaction**: The capability of its members to interact, in some form and to some extent, with other members of the crowd. That is to say, individuals who happen to be at the same time and place with other individuals

but cannot interact with others due to physical or visual/auditory constrains would not be considered members of the crowd.

- **Level of Service**: This concept encompasses multiple aspects of the individual's motion behavior, such as walking speed, ease of mobility, and navigability. This concept was thoroughly explored by Fruin [55] where he describes six levels of service which we present in Table 1.1 along with its corresponding illustrations in Figure 1.2.

- **Locality**: Are individuals required to be at the same physical location to be considered members of a crowd? From a psychological perspective, the answer is 'not necessarily' as first stated by Le Bon [74]. Le Bon illustrates this answer with an example of tragic news broadcasted across a country, evocating a common feeling and disposition to adopt a similar behavior among its citizens.

- **Novelty**: This concerns a crowd's members' ability to behave in a coherent fashion despite their possible unfamiliarity towards the environment and other members, and their lack of proper channels of communication among all its members [106, 127]. This further diminishes the importance of size as a crucial factor to define crowds; for example, an army would not constitute a crowd since their members are well aware of their hierarchy and clear established ways to organize and communicate.

- **Size**: There is no real consensus as to what size constitutes a crowd. Proposing figures is particularly challenging because crowds emerge in different types and environments. However, it is widely accepted to suggest that a crowd should be a sizeable gathering of people [25].

- **Temporality**: Crowds are temporal in the sense that individuals come together at a specific location for a particular purpose and for a measurable amount of time. Again in this respect, there is no consensus as some types of crowds (e.g., sports events) remain assembled for a relatively prolonged period, whereas other types of crowds (e.g., transport station) see members join and leave at a constant pace. However, it is agreed that permanence should be longer than just momentarily for an individual to be considered a member of a crowd.

Level of Service	Description
A	Flow rate of less than 23 people per meter per minute. Virtually unrestricted choice of walking speed. Minimum manoeuvring needed to pass fellow pedestrians. Unrestricted crossing and reverse movements.
B	Flow rate of between 23 and 33 people per meter per minute. Normal walking speeds, restricted only occasionally. Occasional interference in passing fellow pedestrians. Occasional interference in crossing and reverse movements.
C	Flow rate of between 33 and 49 people per meter per minute. Partially restricted walking speeds. Restricted passing movements, but possible with manoeuvring. Restricted crossing and reverse movements, with significant manoeuvring needed to avoid conflict. Reasonably fluid flow.
D	Flow rate of between 49 and 66 people per meter per minute. Restricted and reduced walking speeds. Passing fellow pedestrians rarely possible without conflict. Severely restricted crossing and reverse movements, with multiple conflicts. Momentary flow stoppages possible when critical densities are intermittently reached.
E	Flow rate of between 66 and 82 people per meter per minute. Restricted walking speeds, occasionally reduced to shuffling. Passing fellow pedestrians impossible without conflict. Severely restricted crossing and reverse, with unavoidable conflicts. Flow achieves maximum capacity under pressure, but with frequent interruptions and stoppages.
F	Flow rate variable. Walking speed reduced to shuffling. Passing movements are impossible. Crossing and reverse movements are impossible. Frequent and unavoidable physical contact. Sporadic flow, on the verge of complete breakdown and stoppage.

Table 1.1: Description of Fruin's levels of service.

1.3.3 Types of Crowds

The dynamics exhibited by a crowd depend to a large extent on the type of crowd one is dealing with. As pointed out in [25], A limited amount of research has previously addressed the systematic classification of crowds into types, albeit not to a complete consensus. One widely adopted typology, is the one proposed by Berlonghi in [13] where 11 types are outlined. The work presented in this thesis is applicable to three of these types:

- Ambulatory Crowd: A type of crowd composed of pedestrians whose primary intention is to travel across an environment.
- Panic Crowd: A type of crowd where pedestrians engage in competitive behavior to enter/exit a venue due to a perceived threatening situation.
- Queuing Crowd: A type of crowd where pedestrians display a sequence-like formation to obtain a good/service or to enter/exit a venue.

1.3.4 Working Definition

With a better understanding of the characteristics present in crowds, we come to the realization that not all crowds are created equal, and therefore a relevant distinction is necessary: A crowd that shares a sense of mental unity where a common motivation is shared, a coherently collective behavior is displayed, and the contagion and amplification of emotions occurs, is denominated a *psychological crowd*. Conglomerates lacking mental unity, as is the case of numerous people in a train station who simply happen to be gathered in the same place but for different reasons, are referred as *physical crowds* [66]. Physical crowds may potentially turn into psychological crowds, dictating a different pattern in the dynamics of the crowd. However, having mental unity or not, people present in the same environment are prompt to be influenced by others, triggering emotional responses to cope with the context they experience, and the research presented here is generalized to account for both types. On a final remark, the methodology proposed in this work applies only to crowds that exhibit ambulatory; this includes casual crowds, queues, aggressive mobs and panic crowds. The working definition employed throughout this thesis is presented as follow:

Definition 1. *Crowd: A sizeable group of people gathered in the same place and with enough proximity among its members to convey a sense of togetherness. Mem-*

bers express collectivity (e.g. common goals, interests, sentiments) and behave coherently despite its possible unfamiliarity to the environment or other members. Individuals of this crowd are capable of interacting with other members to some extent, display an ambulatory behavior that is primarily non-static, and retain membership for a period of time longer than just momentarily.

1.4 Emotion

This term came into circulation around 1570 in the French language as *esmotion*, derived from the Latin word *emovere* with the original meaning of 'to set in motion'. In its contemporary usage, it is employed to denote several human affects and is frequently interchanged with feelings, moods and personality traits. On an academic context, this term is much more well defined, although a widely accepted consensus is yet to be achieved. As follow, we present its most commonly accepted features, predominant theories of emotion, measurement approaches, and a working definition to be used in the context of this work.

1.4.1 Features

In the ongoing debate about what an emotion is, several features are agreed upon by many prominent figures in the field of Psychology and Neuroscience.

- Emotions pertain the appraisal of stimuli: Emotions are evoked by events that are relevant to the well-being of a person. In a continuous evaluation of the surrounding environment as well as of internal factors (needs, motivations, values, etc.), emotions aid in producing an adequate behavioral response to cope with the present situation.
- Emotions are driven by motivations: Events that trigger emotions frequently require actions to be taken, thus interrupting an ongoing behavior as the priority in our motivations change. This dynamic change in motivations driven by the emotion evoked helps produce a state of action readiness useful in adapting to relevant events.
- Emotions affect the whole person: Given the importance of the event, we are prompted to take action, for which a series of preparations take place. Consequently, our attention and appraisal are redirected, and systems in our body, such as motor and somatovisceral are tuned in anticipation for behavioral responses.

- Emotions demand priority: Emotions synchronously claim control on our state of readiness and body systems, limiting our attention and awareness to the matter at hand.

1.4.2 Theories of Emotion

In the field of psychology, several theories and models have been proposed aiming to describe the mechanism by which human emotions arise in the brain. Among the most prominent contributions, there are three predominant groups of emotional theories [114] commonly discussed:

Discrete Emotion Theories

Developed from the work of Paul Ekman [46] where he identified six basic emotions (anger, disgust, fear, happiness, sadness, and surprise). These emotions are said to be universal, or discrete, as they appear to be innate to us all, and have an essential role in procuring our survival. Although it is acknowledged that a higher number of human emotions can be identified, these basic emotions are thought to be the building blocks for more complex emotions. Criticism against this theory lies mainly in the lack of empirical evidence from a neurological perspective, where conducted studies have failed to clearly depict dedicated systems for these discrete emotions in the human brain. Although, Ekman himself and other supporters have shifted their views over time, this theory remains relevant in several fields of inquiry.

Dimensional Theories

A family of theories that gained traction due to the contributions of James Russell among other authors [109]. This theory proposes that emotions emerge in a continuum of one or more dimensions, with the classical view on two axes; valence for capturing the goodness or averseness of the emotion, and arousal used to describe its intensity. Under this description, emotions can be represented at any level of valence-arousal or a neutral level in one or both of these dimensions. Theories in this school of thought suggest that a common neurophysiological system is responsible for all the emotions we experience. Some of the major arguments opposing this view are concerned with the universality of emotions, particularly from ethnographic and cross-cultural studies that evidence discrepancies in the way emotions are conveyed and regulated.

Appraisal Theories

Pioneered by Magda Arnold and Richard Lazarus, these theories follow the idea that an appraisal, the process of assessing the significance of an event, triggers an emotional process involving appropriate physiological responses and the conscious experience of emotions. A characterization of this nature implies that emotions are determined by the subjective appraisal of a stimulus, where two types of appraisals are identified, a primary appraisal determining the relevance of a stimulus and a secondary appraisal devising a way to cope with the potential consequences. Unlike the two previous theories, appraisal theories do account for differences in people's emotional responses to the same situation. The main drawback for these theories pertains the actual implementation of appraisal mechanisms in the brain, where no conclusive evidence has been found of the ties between neural substrates and specific appraisal components [90].

In the context of crowds, where we analyze ambulatory movement patterns of pedestrians to understand the underlying emotions of people and the crowd, the family of discrete emotion theories poses a convenient and straightforward representation due to its discrete set of emotions, but the behavioral expression of these emotions in crowds may appear undistinguishable, causing ambiguity in the inference of emotions. Appraisal theories place important attention to the subjective cognitive process of assessing stimuli. However, this process is unobservable from a visual perspective, for which the necessary measurements are not available in crowded environments. Dimensional theories with measurements of valence and arousal in a continuoum appear to be a better match as several studies show the viability to identify these dimensions from visually observed behaviors [147] [17].

1.4.3 Measurement

Under the dimensional theories of emotions, measurements of valence and arousal allow us to produce an inference about the emotions experienced. Due to the limited information that can be extracted from ambulatory patterns, the approach presented here is limited to estimate the valence as related to the motivations and expectations driving pedestrian's actions. Arousal is not accounted for in this method due to the inexactness of its manifestation. As a consequence, the representation of emotion used in this thesis encompasses only the valence dimension. Furthermore, emotions are described at two levels of abstraction, pedestrian and crowd.

Pedestrian Emotion

At the microscopic level, we look at emotions experienced by single pedestrians, where their affective state is characterized by a measurement in the valence axis. In the first iteration of the pedestrian model presented in chapter 5, the emotion of the pedestrian is described by one of three possible discrete states (positive, neutral, negative), as opposed to a continuous scale in its formal definition, due to the constraints in the model presented. In chapter 6, we extend the model's capability to account for valence as a continuous value in the range of 0 for negative, 0.5 for neutral, and 1 for positive valence.

Crowd Emotion

At a macroscopic level, we present a description of emotion that is representative of the emotions experienced by the pedestrians present in a specific subregion of the observed environment, characterized by a measurement in the valence axis. The first iteration of the crowd model in chapter 7 focuses on modelling behavior and motivations, leaving to the second iteration in chapter 8 to address the inference of crowd emotions, measured by a continuous value in the range of 0 for negative, 0.5 for neutral, and 1 for positive valence.

1.4.4 Working Definition

Despite a lack of consensus among leading researchers, Scherer in [114] takes on the task of blending agreed upon commonalities to offer a definition consistent to the existing body of literature. We borrow this definition as part of the theoretical foundation for the work presented in this thesis:

Definition 2. *Emotion: A processes focused on specific events (and thus always have an object); it involves the appraisal of intrinsic features of these objects or events, of their conduciveness with respect to specific need or goals and of their compatibility with norms and values; it affects most or all bodily subsystems in a coherent fashion leading to an integrated mental representation of an episodic emotional quality; is subject to rapid changes due to the constant unfolding of many types of events and the resulting reappraisals of the potential consequences (which in turn change the response pattern); and it has a substantial impact on behavior due to the generation of action readiness (although the actual behavior is also strongly determined by other factors, e.g., situational constraints) [114].*

It is essential to mention that throughout this entire research, we employ the above definition of emotion; however, the emotion will be measured in two cases: pedestrian emotion and crowd emotion. This decision to account for both pedestrian and crowd emotions is intended to provide a more comprehensive view of the observed crowd, while at the same time, both the pedestrian and crowd models are designed to preserve consistency with one another. A dualistic view of the crowd leaves the door open for further exploration into several related phenomena, like contagion and regulation of emotions across members of a crowd, and the interaction and influence between single pedestrians and the crowd they conform. Additionally, it expands the adaptability of our research to a broader range of environments and situations.

1.5 Pedestrians and Crowds

Throughout the content presented in this thesis, we will refer to the pedestrian and the crowd (or sub-crowd) as separate entities that share common elements. These elements are analogous from one another but are defined in a different way. To help in this clarification, table 1.2 provides a list of terms with their corresponding definitions for pedestrians and crowd.

1.6 Research Questions and Objectives

The research presented in this thesis aims to incorporate knowledge in the fields of emotional theories, crowd psychology and crowd behavior analysis to enrich the literature addressing the analysis of behavior and emotion estimation in crowded environments. The main research question in this thesis is:

> **Main Research Question:**
> *How to design a model capable to infer emotions of people in crowded environments?*

We begin to address this question in chapter 2 by a careful examination of affective cues used in the existing literature as well as models intended to analyze crowds. Estimation of human emotions have been successfully addressed using physiological measurements collected from specialized sensors that are not yet readily available in wearable devices that could enable the collection of data in crowded environments as shown in [133]. A second alternative is the observation of facial expressions and body postures which can be extracted

from widely available surveillance cameras, however, visual cues of this nature are frequently occluded in crowds due to high density of people or the position of pedestrians with respect to the surveillance camera. A suitable alternative is the use of the pedestrian motion patterns as we count with existing methods capable to detect and track pedestrian with high accuracy. In this direction, we have devised two main objectives: firstly, to produce a method capable to infer the emotions of individual pedestrians in a crowded environment from observing their ambulatory movement; and secondly, to provide a method to infer the emotions of pedestrians in a collective fashion, also from observing their ambulatory behavior, but with special care to preserve consistency between individual and collective inference of emotions.

Research Question 1.1:
How to model the individual pedestrians in a way that allows to associate their walking movement to their underlying emotions?

This question is explored in chapters 5 and 6 where we propose a pedestrian model based on a hierarchical Bayesian network capable to learn ambulatory patterns to foresee the motivations and expectations of individual pedestrians that allow us to infer their emotion according to the context of the situation. More specifically, in chapter 5 we conduct an experiment where we test our pedestrian model and its capability to predict short term movements as well as its intended destination. Subsequently in chapter 6 we conduct a second experiment where we extend the pedestrian model to formally incorporate the elements of motivations, expectations and the way in which these relate to the inferred emotion.

Research Question 1.2:
How to produce an inference of the collective emotion of pedestrians in a way that is consistent with their individual emotions?

In chapters 7 and 8 we present the crowd model, sharing a hierarchical Bayesian network similar to that of pedestrian model capable to capture the elements of motivations and expectations in an aggregated way to produce an estimation of the emotion of pedestrians found in close proximity. Approaching the estimation of emotions at two levels of abstractions provides a more comprehensive view of the dynamics of a crowd, and enable us to investigate further the interaction between single pedestrians and the crowd they belong. Additionally, considering this research relies on surveillance cameras to observe the crowd, a

dualistic way to view the crowd helps to cope with different environmental challenges. For instance, pedestrian detection & tracking algorithms function well in crowds with low density and minimal occlusion, but their performance decays rapidly as the number of people increases, causing the detected trajectories to be fragmented. To deal with high-density crowds, we apply crowd counting algorithms that provide an estimation of the number of people in the scene. Taking advantage of these two types of algorithms, the pedestrian model is better suited for sparse crowds where unfractured trajectories can be extracted, while the crowd model is less susceptible to the density level because it requires only partial trajectories and counting estimations.

1.7 Research Approach

In this thesis, we propose a method capable to analyze the walking trajectories of pedestrians present in crowded environments to produce an estimation of their corresponding emotions. The proposed method consist of two models: a pedestrian model focused on delivering a estimation of individual pedestrian's emotions, and a crowd model centered on estimating the emotion experienced by multiple pedestrians in close proximity. For both the pedestrian and crowd model, an experiment is conducted on each stage of development to test its validity and accuracy.

Concerning the selection of a suitable approach to observe the crowd, three options were evaluated: physiological measurement, facial and body expressions, and walking trajectories [129]. Sensors capable to obtain physiological measurements for inferring emotions have a good record in the existing literature but for this case were dimmed impractical, firstly due to the lack of publicly available datasets capturing physiological measurements in crowded settings; secondly because collecting data would require careful calibration of sensors, add cost and increase complexity in preparing experiments with a significant number of participants; and finally, ethical principles would limit the naturalness and the potential scenarios to explore. Facial and body expressions can be captured with widely available surveillance cameras, however the computational cost increases rapidly when dealing with medium to large crowds; additionally, faces and bodies can be occluded depending on the position of pedestrians and density levels, hence this option was also judged not suitable. Walking trajectories can be easily extracted via surveillance cameras, preserving the naturalness of the observed situation; on the downside, walking trajectories introduce higher ambiguity in its relation to the emotion experienced by the

pedestrians. After a careful consideration of the above mentioned choices, this method opted to use surveillance cameras to observe the walking trajectories of pedestrians. Selecting this approach facilitates the acquisition of data via currently available and robust techniques developed in the fields of computer vision and crowd behavior analysis [59]. For example, algorithms for pedestrian detection can be used to extract pedestrian trajectories, and crowd density estimation algorithms provide an approximate measure on the number of pedestrians present in the scene. In practice, the experiments conducted in this thesis made use of annotated datasets [144] and social-force based simulation models [64] [115].

Once the walking trajectories of pedestrians are collected, we proceed to incorporate the theoretical foundation to formulate a hypothesis on the association between walking trajectories and emotions. We start with the assumption that observable behavior, walking trajectories in this case, bear a relationship with the pedestrian's mental and emotional states, as supported by behaviorism [22]. Employing field theory [20], a psychological theory examining patterns of interaction between people and their environment, we propose that pedestrians present in an environment walk in accordance to a motivation, i.e., their desire to reach a destination. Expectation then comes into play as an attempt to foresee the effort required to meet this motivation and guide our decision making process [97]. An intricate factor in the cognitive process of decision-making is emotions, aiding to evaluate the perceived stimuli to produce a behavioral response [32] [11] [120]. The theory of affect heuristics further supports the central role of emotions, as it proposes that emotions serve as mental shortcuts to make decisions quickly and efficiently [52]. Emotions are measured in a continuous-valued valence axis consistent with dimensional theories of emotion [109]. With the theoretical foundation presented in the previous lines, our working hypothesis states that walking trajectories result from the conditional dependence of motivations, expectations, and emotions.

Having identified the key components (i.e., walking trajectories, motivations, expectations, and emotions), we proceed to design a computational model using Bayesian networks as we find this a suitable method to our research questions. Bayesian networks [93] are a type of probabilistic graphical model that represents conditional dependence among random variables to efficiently produce inference about other random variable. A variation of this model called dynamic Bayesian networks are capable to describe the relationship among random variables over time, as in the case in this work. Furthermore, this method accommodates the use of continuous-valued random variables like the sequence of observations of a pedestrian in a walking trajectory, and discrete-valued vari-

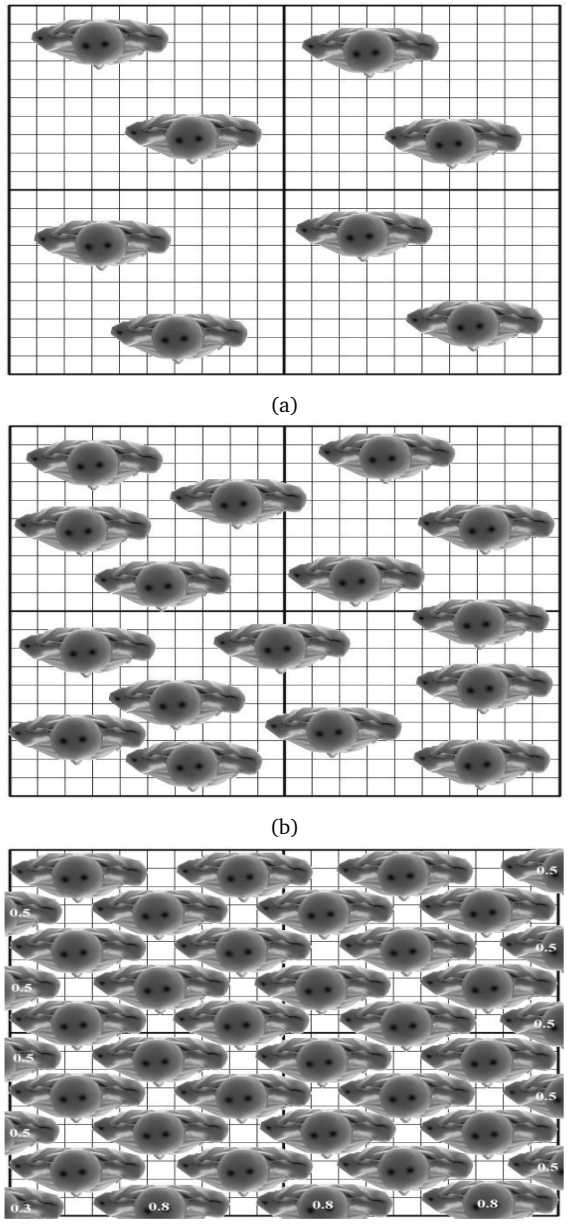
ables like the motivation of a pedestrian guiding its movement. Since our models require the use of multiple random variables organized in a hierarchical way, the approach employed is denominated hierarchical dynamic Bayesian networks HDBN. Two HDBN are developed throughout this thesis, one to describe the pedestrian as a single entity and another one for the crowd to describe multiple pedestrians.

A series of experiments were conducted at different stages of development for the pedestrian and crowd model to assess their validity and accuracy. Chapter 5 presents the first iteration of the pedestrian model and an experiment is conducted employing a simulation tool to produce a virtual crowded environment where pedestrians interact in a complex environment to reach a particular destination. In this first experiment, our model is evaluated for its predictive capability to infer the desired destination (i.e., motivation) and short term movement of individual pedestrians. In chapter 6, the elements of expectation and emotion are incorporated; and an experiment combining both real-world and simulated datasets is conducted to confirm the model's capability to infer pedestrian emotions. Our first version of the crowd model is introduced in chapter 7 where an experiment using a real-world dataset evaluates the model's ability to describe the behavior of multiple pedestrians as a single entity. Finally in chapter 8, an improved version of the crowd model is provided, accompanied by an experiment that tests our model in a wide variety of scenarios.

1.8 Research Contributions

The work presented in this thesis makes several contributions to the fields of crowd behavior analysis and affective computing. First, it provides a pedestrian model based on hierarchical Bayesian networks to describe the ambulatory behavior of pedestrians in a way that accounts for the motivations and expectations driving their actions. This same model is further expanded to infer the emotion of pedestrians in a continuum valence axis, going beyond abnormality detection or simple behavior classification. Secondly, it proposes a crowd model, also based on hierarchical Bayesian networks and with a structure equivalent to that of the pedestrian model, capable to capture the collective ambulatory behavior of pedestrians, inferring aggregated representations of group motivations and expectations to produce an estimation of their collective emotion in a way that is consistent to that of their individual emotions. Furthermore, we employ a data-driven approach, enabling the pedestrian and crowd models to learn in unsupervised way and adapt to multiple contexts where ambulatory

crowds are present. Finally, these models are developed with a solid foundation in psychology literature, an aspect frequently neglected in computational models.



(a)

(b)

(c)

Figure 1.1: Three levels of crowd density. (a)Low Density: 20 people per 10 square meters.(b)Medium Density: 40 people per 10 square meters. (c)High Density: 84 people per 10 square meters. (Pictures taken from http://www.crowddynamics.com/ Myriad%20II/ Anthropomorphic.htm)

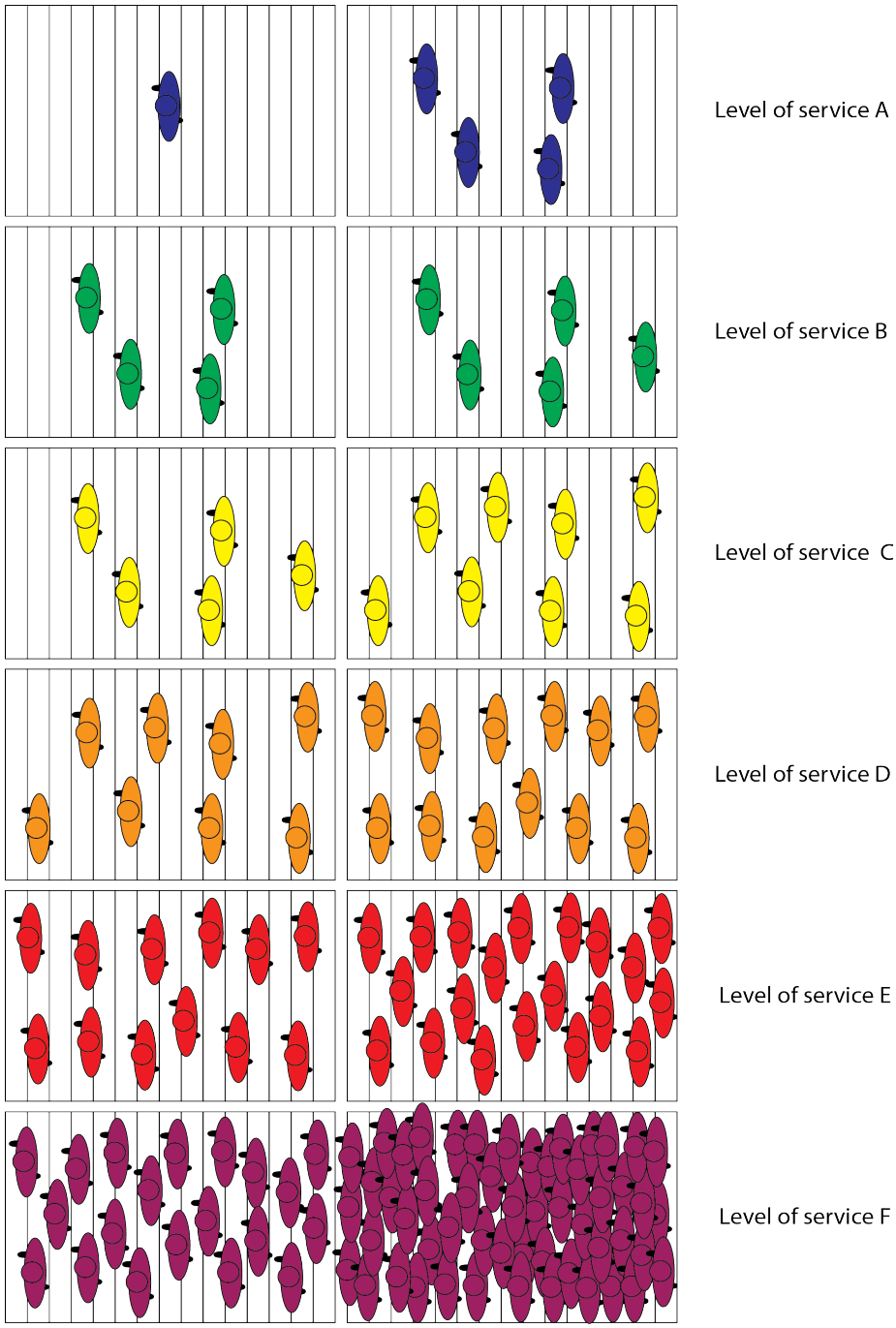


Figure 1.2: Illustration of Fruin's levels of service.

	Pedestrian	Crowd
Behavior	A sequence of locomotive actions to move from one place to another in response to a particular motivation or due to interactions with the environment.	A sequence of discrete state transitions over time with different transition sequences corresponding to a particular behavior.
Emotion	A pedestrian's affective state described by a discrete or continuous value in a valence spectrum based on the multidimensional theory of emotions.	The affective state representative of the emotion experienced by the pedestrian members of a crowd or sub-crowd, described by a discrete or continuous value in a valence spectrum based on the multidimensional theory of emotions.
Expectation	A pedestrian's estimation of the effort in terms of distance and/or time required to reach the destination associated with its motivation.	A prediction of the expected sequence of state transitions to be observed in a crowd or sub-crowd.
Motivation	The reason for a pedestrian to engage in a particular walking trajectory and interact with an environment to achieve an objective.	The collective intention of a crowd or sub-crowd to follow a particular pattern of state transitions.
State	The physical location or sub-region of the environment a pedestrian is currently located in.	A particular configuration of the feature vector used to represent the crowd.

Table 1.2: List of definitions for entities pedestrian and crowd.

Chapter 2

Literature Review

The matter of emotion recognition and human behavior analysis has gained a modest interest over the past years, with research conducted in the fields of neuroscience, general psychology, and computer sciences. The main focus has concerned the recognition of prototypical expressions, employing the discrete emotion model in most cases, based on data collected in controlled settings. On a more recent trend, the interest has shifted towards recognition of emotional displays observed in real-world settings where subtle, continuous, and context-specific interpretations of affective displays can be captured, and in which multiple modalities for analysis and recognition of human emotion can be applied [7] [61]. This chapter examines the existing body of literature concerning the estimation of emotions in crowded environments, addressing three main aspects: crowd modeling techniques, emotion recognition methods, and relevant datasets.

2.1 Crowd Modeling Techniques

One of the first concerns in the analysis of crowds pertains how to describe a crowd, cascading into several sub-tasks: detection, localization, and tracking of pedestrians in crowds. Once developed a proper model of the crowd, subsequent behavior analysis can be performed in a more reliable and accurate way. Broadly speaking, we can group the existing crowd modeling techniques as follow: motion-based techniques, appearance-based techniques, social force models, and simulation models. Another relevant distinction to be made is between

data-driven and model-based approaches. Data-driven approaches rely on machine learning methods to construct a model and require significant amounts of data; the more available data, the better the model will generalize. Model-based approaches rely on a deep understanding of the process of interest, benefiting from scientifically established relationships; however, models of this type can't accommodate infinite complexity and usually must be simplified. Appearance based techniques are data-drive, whereas motion based techniques and social force models are model-based approaches. Given the scope of this thesis concerns the inference emotions in real-world crowds, we do not explore methods for simulating emotion-based behaviors in crowds. Finally, as a general assumption, we presuppose the input signals to be visual, with some exceptions as is the case of social force models.

2.1.1 Motion Based Techniques

Approaches of this nature aim to model the crowd by focusing on the movement caused by pedestrians and their interactions, and the movement can be described at either macroscopic or microscopic levels. Modeling macroscopic motions is mainly done with the transformation of movement parameters estimation [18]. At the microscopic level, estimating the motion of single individuals is often used for detecting and tracking pedestrians [146]. Techniques of this nature are reliable and relatively fast but they presuppose the crowd is in constant motion, limiting their applicability to ambulatory crowds. Another important consideration is that environmental conditions such as illumination, camera stability, and occlusions, significantly affect their performance. Algorithms that rely on background subtraction are usually employed in motion-based methods for pedestrian detection and crowd behavior modeling. Focusing on microscopic movement, multiple background subtraction methods have been developed based on Gaussian Mixture Models (GMM) [125], namely [26] and [138], among others. Others have explored the use of Average of Gaussian's such as [37] and [148]. Optical flow is a widely employed approach in the scope of macroscopic movement, with frequent use in estimating the motion of crowds for behavior analysis, with a particular interest in group behaviors.

2.1.2 Appearance Based Techniques

Techniques in this realm are distinctive due to their capability to detect visual appearance features, an aspect absent in motion based methods. Focus-

ing on appearance enables the distinction between motion exerted by pedestrians and motion caused by other factors like camera movement. In contrast to motion based techniques suited for ambulatory crowds, techniques centered in appearance are highly suitable for behavior analysis in stationary crowds where little to no motion is observed. On the downside, appearance based techniques must undergo a more extensive training phase with a higher volume of data in similar environments to yield accurate results [103] [123] [27]. In crowd detection, several features and descriptors are employed, for instance, histogram of oriented gradients (HOG) [38], where occurrences of gradient orientation are counted at specific regions of an image. This is sometimes used in combination with other descriptors, particularly histogram of optical flow (HOF) [121], histogram of oriented tracklets (HOT) [92], and color triplet comparison (CTC) [75]. When dealing with larger crowds, where density is high, body-part detectors can be of great help because the performance of full-body detectors deteriorates rapidly due to the occlusion caused by overlapping pedestrians; in these cases, upper body-part detectors [146] tend to perform better. Lastly, with the advent of machine learning, some methods have been applied to crowd behavior analysis with great success, e.g., support vector machine (SVM) [35] for detecting and modeling the appearance of a crowd.

2.1.3 Social Force Models

The social force model (SFM) developed by [65] posed a significant advance in the study of crowd behaviors from a sociological perspective. This agent-based model describes the behavior of people in crowds as a result of attractive and repulsive forces experienced by people in social settings, also influenced by constraints in the environment and personal motivations. This model also includes contextual and social aspects introduced by the individual's interaction with other people and objects. The resulting force experienced by each individual under the social force model is:

$$F^k = F_d^k + \sum_{h \neq k} F_r^{kh} + \sum_A F_a^{kA} + \sum_Q F_r^{kQ} + \xi^k \tag{2.1}$$

where F_d^k represents the motion force of every individual guided by their intended destination. A repulsive force F_r^{kh} is considered due to the individual's tendency to mantain a distance from other people to avoid collisions. Another repulsive force F_r^{kQ} is included to capture the individual's intention to avoid physical obstacles present in the environment. Conversely, an attractive

force F_a^{kA} is accounted for due to individual's attraction towards other people or points of interest. To conclude, the variable ξ^k is included to account for unexpected movements of individuals not captured in the previous forces.

2.2 Emotion Recognition Methods

Inspired on neural mechanisms revealed by the recent work of Damasio [39], the authors of [67] and [15] propose ASCRIBE, an agent-based model to describe the interplay of mental states (*emotions*, *beliefs* and *intentions*) of individuals in the decision-making process under stressful situations. ASCRIBE is defined as having an external and internal level of operation; at the external level it incorporates mechanisms for mirroring mental states between individuals, at the internal level it describes how emotions and beliefs affect each other and how both affect a person's intentions. The model was put to the test by simulations and an empirical study case that compared four models and showed the ASCRIBE model to yield higher prediction accuracy. Expanding the concepts applied in the ASCRIBE model, the multi-agent-based model presented in [14] formalizes several concepts of emotion contagion spirals based on fundamental aspects at the individual level: the senders current emotional state and the extent to which the sender expresses the emotion; the strength of the communication channel from sender to receiver; the receiver current emotional state, its openness or sensitivity for the received emotion, bias to adapt emotions upward or downward and tendency to amplify emotions. Although no empiric validation is provided, the model is tested with simulations and mathematical analysis, and it produced interesting emerging patterns identified in psychology literature such as the upward and downward emotion spirals. Further studying the role of emotions in the decision-making process under stressful situations, and with similar concepts to ASCRIBE, an adaptive agent model for affective social decision making is proposed by Manzoor et al in [119] and later extended in [83] to account for emotion regulation and contagion. This model incorporates Hebbian learning principles to adapt the agent's decision-making process, but as found in the experiments conducted, it did not yield significant discrimination in the agent's decisions. Regulation is approached by antecedent-focused strategy (regulation before an emotional response has an effect on behavior), modeling it as a dynamic interaction between internal mental states and contagion is implemented as described in [14].

Focusing on the task of emotion estimation, the framework introduced in [91] addressed the recognition of individuals' membership and emotions within

a group setting by means of multi-modal analysis of facial and body expressions. Faces are represented by facial landmark trajectories and extended volume Quantised Local Zernike Moments (QLZM) [112], and encoded into Fisher Vector (FV) [111] representations as input to a GMM classifier to recognize emotions in arousal and valence dimensions. The framework was tested with a self-collected dataset of 3 groups of 4 individuals each, monitored while watching a movie. The proposed approach vQLZM-FV outperformed the compared methods, namely Facial landmarks, body HOG and body HOF. Although this approach focused on investigating the affective response of individuals while watching long-term videos, it is theoretically applicable to crowds under the assumptions that crowd members' face and body are visible for a long-enough interval and within an acceptable resolution.

The authors of [6] summarize their previous work on three bio-inspired probabilistic algorithms for perception of emotions from crowd dynamics. The first algorithm starts by partitioning the environment using an Instantaneous Topological Map (ITM) and a Dynamic Bayesian Network (DBN) is employed to model conditional interactions occurred in each sub-region. these interactions are then converted into super states using a Self-organizing map (SOM) and the occurrence of these super states (events) are encoded by a Gaussian mixture model as positive or negative emotions. The second algorithm starts with the detection of events, collecting them over time to obtain behavioral patterns which are then clustered into classes by means of a DBN; the distribution of these classes are modeled using GMM, building one model for positive and one for negative emotions, to finally detect the emotional state by a likelihood ratio test. In the third algorithm the trajectory of single individuals are expressed as transitions (events) between sub-regions using a DBN and separated models are constructed for the event sequences labeled with a positive or negative emotion, to conclude with a log-likelihood test to determine the emotion according to the movement pattern exhibit by the individual. All three approaches are tested under a simulated scenario, showing the third algorithm to yield the highest emotion prediction accuracy according to the experiments conducted.

2.2.1 Methods Comparison

A quantitative or qualitative comparison of methods intended for emotion estimation in crowds remains a challenging task due to the lack of a common dataset with appropriate annotations of pedestrian motion, behavior and emotions in a sufficient variety of scenarios, as it will become evident in the following section. Another difficulty in comparing methods, is the framing of the

problem, scenarios accounted for, and specific aspects to address. This section attempts to provide a fair comparison between methods by discussing and contrasting the aspects they cover. Again, it is important to mention that we only consider methods that address the recognition of emotions in physical crowds, hence methods focused on simulated and online crowds are not included. Several methods applicable for emotion recognition in crowded environments are listed in table 2.1 along with four key aspects used for comparison. This list of methods is not exhaustive but rather representative of the relevant body of literature.

The first aspect of comparison, types of crowds allows to establish the context under which the method is applicable, which is a relevant point since no current method is suitable for all types of crowds. The methods [85] [91] [7] are appropriate only for crowds where pedestrians remain in a relative fix position. On the opposite extreme, the methods [91] [6] strictly require pedestrians to be in motion. The work in [16] is a particular exception where the authors study the effect of emotions in crowds as a response of transitioning between fixed to panic behavior. Our method is suited for crowds with ample movement as observed in ambulatory and panic crowds, as well as crowds with limited movement such as queuing crowds.

The inference of emotions in crowded environments is performed at the individual and/or collective level. The methods in [91] and [8] study the inference of individual emotions irrespective of their interaction with other people. [6] provides individual emotion estimations resulting from the interactions with other pedestrians and the environment. [16] also accounts for individual emotions with a strong focus on emotion contagion and regulation. [85] and [7] specifically deal with inferring collective emotions evoked from the interaction between members of a group. Our methods opts for a dualistic approach were the emotions are estimated at both individual and collective levels.

The choice of behavioral cue employed to infer emotions is an important one as it directly impacts on the applicability of the method. For instance, [7] depends on the observation of faces, which is computationally expensive when dealing with large crowds, or even unfeasible under low image resolution or constant occlusion. The use of body postures in [85], [91], and [8] is better suited than facial expressions in crowded environments with the caveat that portions of the body may also be highly occluded in medium to large crowds. Walking trajectories are employed in [16], [6], and our method, which can be extracted with relative ease employing people detection and tracking algorithms, with minimal impact in accuracy even in crowds with high density; however, this choice of behavioral cue conveys less emotional significance in

comparison with facial and body expressions.

Regarding the actual emotions each method is attempting to estimate, the methods [16], [85], [8], [7], and [6] target a varying list of discrete emotions based on the discrete emotion theories, whereas [91] applies a continuous-valued scales of arousal and valence based on the dimensional theories of emotion. Similar to [91], our method uses a continuous-valued scale on a valence axis as it is less restrictive when addressing the lack of consensus in collective emotions, and we neglect the arousal component due to its ambiguous association to walking trajectories.

Finally, in terms of quantitative performance of emotional inference, a comparison is difficult due to the different scope, dataset, metrics and experimental settings adopted by each method we evaluate, however, as follow we summarize some key results. [16] conducted experiments based on a real-world panic situation where two variations of their model were tested; the variation of the model neglecting emotion contagion reported a 0.66 average error rate in detecting fear, while the model accounting for emotion contagion produced an average error rate of 0.54, providing evidence to the importance of emotion contagion in panic crowds. The work in [85] produced a series of simulated scenarios where pedestrians displayed contrasting body postures, an average of 87% of the emotions were correctly identified in a varying size of crowds. The method presented by [91] evaluated their model employing footage of a small number of participants seated in front of a camera; arousal was inferred with a minimal MSE of 0.013 and std of 0.03, and valence with a minimal MSE of 0.014 and std of 0.03. The study conducted in [8] recorded kinematic data using motion sensors attached to participants while they walk $5m$ in a straight line; the interaction between emotion and several body parts was measured with the strongest and most consistent statistical inference found for fear. The work of [7] employs the Mars-500 dataset where footage of isolated astronauts captures their interactions while playing a CT game; their method allowed to find different response functions that identify events evoking an emotional response, with the best performance reaching a aggregated score of 13.46 indicating how well the functions explain the observed dynamics of facial expressions. [6] evaluates their model to identify emotions on walking trajectories generated via a simulation tool, with positive emotions detected with a 94.54% accuracy and negative emotions with an accuracy of 89.63%. Our approach is presented in the following chapters with several experiments conducted over each iteration of our models. In their final iteration, our pedestrian model for inference of individual emotions achieves a minimal MSE of 0.015 whereas our crowd model for collective emotions reaches an average MSE of 0.02. A more extensive dis-

cussions of our results is presented in the remainder of this thesis.

2.3 Datasets

A common test bed is essential to measure and compare performance among different methods. Within the scope of our research, this section examines two types of datasets, those designed for affective states recognition and those intended for crowd analysis tasks. The objective is to determine whether the reviewed datasets are suited to test computational models dealing with emotion estimation in crowds. The datasets considered in this section are not exhaustive but rather representative of the diversity available for these tasks.

2.3.1 Affective Datasets

This subsection considers datasets intended for affective-related tasks using different types of sensors and annotation formats, not limiting to those fitted for crowds, to provide a comprehensive view of the available options.

Delving in the task of detecting emotional states, facial expressions have become a popular choice due to their universality and intrinsic relation to emotions [47]. By means of conventional cameras and in a controlled environment, the datasets CMU [89] [40] and FER-2013 [57] collected static images of facial expressions from participants who were requested to act different emotional states following the discrete emotions scheme [46]. CMU data consists of black and white face images of individuals taken in different poses (straight, left, right, up), expression (neutral, happy, sad, angry), eyes (wearing sunglasses or not), and size. Similarly, FER-2013 focuses among other tasks, on the facial expression recognition task, providing portrait images of single participants acting expressions of happiness, sadness, anger, surprise, and neutral. Aiming to simplify the collection of static images and to reach a greater number of participants, the authors of the Gamo dataset [78] made use of a web-based interface were participants play a game by performing specific facial expressions captured by a web camera. Each static image captures only one person, focusing on the facial area. Progressing from static images only, the CK+ dataset [82] provides 22 sequences of images including 27 subjects, where each participant enacts a series of facial expressions which are later annotated in terms of facial action units sequences [49], and emotion labels are revised and validated.

Expanding the scope of behavioral markers, the dataset CREMA-D [23] contains short videos of individual participants displaying facial and vocal expres-

sions for the study of multi-modal emotion expression and perception, whereas the dataset LIRIS-ACCEDE [10] goes one step further and captures body postures in addition to facial and vocal expressions. In comparison to similar datasets that contain few video resources and limited access due to copyright constraints, LIRIS-ACCEDE consists of 9,800 good quality video excerpts with a large content diversity. Annotations regarding affective states are achieved by crowd-sourcing pair-wise video comparison protocol, helping to ensure the annotations are consistent, as confirmed by a high inter-annotator agreement, despite the diverse background of the annotators. However, as the authors in [30] argue, using conventional 2D cameras lacks robustness as this kind of cameras are subject to poor illumination and changes in lighting conditions. In response, they propose the use of Kinetic cameras as these are able to capture depth, and produce a dataset containing 3D models of several participants performing multiple facial expressions, categorized to a particular emotional label (normal, happy, ad, surprise, angry). Moving from emotions (brief affective states) to moods (long term affective states), the work in [69] introduces the EMMA database which employs both 2D and kinetic cameras, and provides longer intervals of data capture as the dataset is intended for mood recognition. Focusing on physiological measurements, the DEAP dataset presented in [71] [102] collected the electroencephalogram (EEG), electrooculogram (EOG), Galvanic skin response (GSR), blood volume pressure (BVP), temperature and respiration signals of participants. And a frontal video face was recorded for some of those participants. One-minute long excerpts of music videos were used as the stimulus to elicit emotions along the four quadrants of the arousal-valence plane. All the above mentioned affective datasets are listed in table 2.2 along with important characteristics.

In trying to apply any of the examined affective datasets to the task of emotion estimation in crowded environments, several deficiencies become evident:

(a) *Naturalness*: An important aspect of collecting behavioral markers concerns the spontaneity in which these are elicited. When manifesting emotions in an acted or even induced way, there is no validation of the emotional labels as these refer to what was requested rather than what was actually displayed by the participants. For this reason, physiological markers are preferred as they are involuntary; however, the majority of available affective datasets rely on behavioral markers elicited in an acted way.

(b) *Interaction*: Currently, affective datasets focus on the individual, yet no dataset examines the dynamics of emotions in groups under natural set-

tings, how interacting with multiple individuals and other psychological factors influence emotional states.

(c) *Applicability*: The examined datasets rely on facial, vocal and body expressions as well as physiological markers. In crowded environments faces are not always visible, and face-based emotion classifiers become computationally expensive as the number of individuals increases. The use of vocal expressions is subject to environmental noise and it becomes rapidly inefficient with multiple individuals. Depending on the position of the camera and density level, the body of individuals is unlikely to be visible. Physiological markers require several equipment components, rendering this considerably more challenging for crowds.

From the aforementioned points, it becomes evident that the currently available affective datasets are not suited for testing methods intended to estimate emotions in crowded environments where a high number of people is observed.

2.3.2 Crowd Analysis Datasets

The fields of computer vision and crowd analysis are favored with an overgrowing importance and share a common interest in studying crowded environments, resulting in multiple datasets produced. In compiling these datasets, cameras remain the preferred sensor for studying crowds due to the already widespread use of surveillance cameras in most public spaces. Depending on the focus of study, datasets are designed to capture the desired circumstances, for instance, the popular dataset PETS 2009 [51] collected image sequences from multiple cameras with the aim to serve as a test bed for algorithms intended for people counting, density estimation, people tracking, flow analysis and event recognition. All the presented situations are mainly poor in terms of emotional behavior, except for the scenario S3 (event recognition) where an evacuation (rapid dispersion) is observed and can be associated to an emotional state of fear. The authors of CAD [62] recreated several normal collective behaviors adding the challenges of change in illumination and wavering trees in the background, however, the captured situations are not representative of any distinctive emotional behavior. Taking advantage of the large number of people attending the World Exposition of 2010 in Shanghai, the massive dataset Shanghai Expo 10 [142] was gathered. It provided a large amount of annotations at a regional level denoting crowd density, collectiveness and cohesiveness features under normal situations, but it lacks any relevance for inferring affective states. Focusing on groups and crowds, the authors of [116] presented the

Atomic Group Action dataset targeting the dynamics of group formation, yet no meaningful emotional behavior is exhibit. Rabiee's dataset [104] provides some emotional-rich situations such as panic and fight, although in a staged way. Finally, the S-hock dataset [117] focuses on the behavior of spectator crowds with rich annotations at the individual level, enabling the addition of further affective annotations although restricted to this type of stationary crowds. All the above examined datasets are listed in table 2.3 along with relevant features.

Under realistic conditions, video cameras can suffer of poor illumination and changes in light conditions. However, as shown by the examined datasets, conventional video cameras remain to be the most practical alternative to observe crowds. As concluded from the previous inquire on available datasets, those intended for affect tasks are not suited for evaluating methods intended for crowded environments. As for the datasets intended for crowd analysis, only the S-hock presents sufficient relevant data but it is limited to spectator crowds. A well suited dataset for emotion estimation in crowds needs to capture diverse and meaningful behaviors accompanied with well validated emotion annotations, ideally for multiple types of crowds and different emotions.

	Types of Crowds	Type of Emotion	Behavioral Cue	Emotion Labels
T. Bosse et al. [16]	Panic, audience	Individual	Walking trajectories	Discrete: fear
J. McHugh et al. [85]	Audience	Collective	Body postures	Discrete: happiness, sadness, anger, fear.
W. Mou [91]	Audience	Individual	Body postures	Continuous: arousal, valence.
A. Barliya et al [8]	Ambulatory	Individual	Body postures	Discrete: Happiness, sadness, anger, fear.
I. Barakova et al. [7]	Audience	Collective	Facial expressions	Discrete: happiness, sadness, anger, surprise, fear, disgust, neutral.
M. Baig et al [6]	Ambulatory	Individual	Walking trajectories	Discrete: Positive/negative.
Proposed Method	Ambulatory, queuing, panic	Individual, collective	Walking trajectories	Continuous: valence.

Table 2.1: Comparison of methods for emotion recognition applicable to crowded environments.

Dataset	Modality	Sensory Data	Annotations	Naturalness
3D Face Model [30]	Facial expressions	Kinetic camera	Normal, happiness, sadness, surprise, anger	Acted
CK+ [82]	Facial Behavior	Image sequences	Anger, disgust, fear, happiness, sadness, surprise, contempt	Acted
CMU [89] [40]	Facial expressions	Static images	Happiness, sadness, anger, neutral	Acted
CREMA-D [23]	Facial and vocal expressions	Camera	Happiness, sadness, anger, fear, disgust, neutral	Acted
DEAP [71]	Facial expressions, physiological measurements	EEG, EOG, GSR, BVP, temperature, respiration	Valence, arousal, dominance, liking, familiarity	Induced
EMMA [69]	Facial and body expressions	Camera, kinetic camera	Valence, arousal	Induced and acted
FER-2013 [57]	Facial expressions	Static images	Happiness, sadness, anger, surprise, disgust, fear, neutral	Acted
GaMo [78]	Facial expressions	Static Images	Anger, disgust, fear, happiness, neutral, sad, surprise	Acted
LIRIS-ACCEDE [10]	Facial, vocal and body expressions	Camera	Valence, arousal	Acted

Table 2.2: Affective Datasets

Dataset	Modality	Sensory Data	Annotations	Naturalness
Shanghai Expo 10 [142]	Crowd movement	Camera	crowd density, collectiveness and cohesiveness	natural
Rabiee's [104]	Crowd movement	Camera	Panic, fight, congestion, obstacle, neutral behaviors	Acted
PETS 2009 [51]	Crowd movement	Image Sequences	Pedestrians' bounding box and location	Acted
CAD [62]	Crowd movement	Camera	Bottleneck, departure, lane, arch/ring and blocking crowd behaviors	Acted
S-Hock [117]	Individual behavior	Camera	People detection, head detection, head pose, body position, posture, locomotion, action/interaction, supported team, best action, social relation.	Natural
Atomic Group Actions [116]	Group actions	Camera	Group-group actions (formation, dispersal, movement) and group-person actions (person joining, person leaving)	Natural

Table 2.3: Crowd Analysis Datasets

Chapter 3

Emotion Annotation Schemes

In research related to emotions, it is necessary to establish a standard metric by which to represent and annotate emotional states. Despite an unfinished debate over what scheme most accurately captures the essence of emotions [136], the choice of scheme is usually dependent on the research's objective and the signals available for observation. This chapter presents the most commonly used schemes and the one employed in this research.

3.1 Existing Schemes

In the realm of emotion annotation schemes, the four most common choices are basic emotions [48], multidimensional [110], positive/negative, and appraisal schemes [113]. As follow, we present these schemes along with examples of their usage in previous studies.

3.1.1 Basic Emotions Scheme

This scheme employs Ekman's Basic Emotions Theory [48] as its foundation to produce categorical annotations. In this theory, Ekman proposes the existence of six basic emotions universally recognizable by observation of facial expressions. The basic emotions under this theory are anger, disgust, fear, happiness, sadness, and surprise. This scheme has gained popularity over time as other researchers have confirmed Ekman's findings. Due to the nature of the emotion theory associated with this scheme, its use is suitable for categorical annotations

when the observed affective cues are facial expressions as shown by numerous research [1] [124] [139].

3.1.2 Multidimensional Scheme

Derived from the Multidimensional Theory of Emotions [110], this scheme envisions an emotion as a point in a dimensional space, with each dimension corresponding to a characteristic of emotion. The most common features are Valence (level of pleasantness), Dominance (level of potency-control), and Arousal (level of intensity). The effectiveness of each dimension to accurately captures changes in emotional states remains in a debate with proponents such as Fontaine placing higher relevance to Dominance. This type of scheme is widely used in research focused on physiological measurements, as well as behavioral responses. For example, in the dataset DEAP [71], measurements of electrical activity in the brain (EEG) were highly correlated with reported valence, whereas skin conductivity (GSR) was a reliable indicator of arousal. Datasets focusing in visual affective cues have also employed this scheme, that is the case of HUMAINE [43], SEMAINA [86], and IEMOCAP [21], where speech and facial expressions along with body postures are observed to produce annotations in a multidimensional affective space.

3.1.3 Positive/Negative Scheme

This scheme is a simplification derived from a multidimensional theories of emotions where only one dimension (valence) is accounted for, with two discrete states (positive and negative), with an third state of neutral in some studies when there is not enough confidence to produce a positive or negative label. The need for a scheme with this level of simplicity arises when the source of information lacks relevant affective cues to make a more informative assessment of the underlying sentiment. This is particularly well illustrated in research related to natural language processing [96] [36] [99], where the source of information is text-based therefore providing hints by the choice of words but lacking the nuances conveyed by facial expressions, body language, and the pitch and intonation of the voice. A variation of this scheme is often employed in crowd behavior analysis where the main objective is to classify the displayed behavior of a crowd as normal or abnormal [58] [88] [135]. One shortcoming of this scheme is the absence of labels that reflect differences between mild and strong agreement towards a positive or negative state.

3.1.4 Appraisal Scheme

This scheme is based on the appraisal theory of emotion [113] where the elicited emotion is not only yield from a particular stimulus, but also accounts for contextual and environmental changes such as novelty, pleasantness, goal-based significance, coping potential and compatibility with standards. In short, it describes the evaluation of the manifested emotional expression as a continuous and changing process of appraisal and reappraisal in time. The work presented by Barakova et al. in [7] employs this scheme where they analyze video recordings of facial expressions during collaborative interactions in which time-dependent components of facial expressions are extracted and interpreted by a mathematical model of emotional events to find locations, types, and intensities of the corresponding emotional events.

3.2 Proposed Scheme

3.2.1 Walking Trajectories as Affective Cues

Previous studies [71] [43] show physiological measures to yield more consistent results over behavioral responses. However, in the study of naturally occurring crowds, capturing physiological measurements remains still an impractical task due to reasons discussed in section 1.7. On the other hand, most infrastructures and spaces intended for crowds are already equipped with surveillance cameras, and for this reason, the acquisition and use of behavioral cues appear to be more plausible. However, facial and body expressions aren't always observable due to temporal occlusion, high density, or the walking direction of the pedestrians. In the presence of these restrictions, we have opted to use the walking trajectories of pedestrians as the behavioral cue over which emotions are inferred and annotated for this research. This is a reasonable approach, as evidenced by previous studies [50] [8] [101]. A walking trajectory is presented as a time series of the positions of a pedestrian over a period of the observation. Considering the available affective cues, a multidimensional scheme is employed where only the valence-axis is considered, in a continuous-valued scale from 0 (negative) to 1 (positive).

3.2.2 Mapping Walking Trajectories to Labels

To construct an association between walking trajectories and emotional valence, we focus on environments intended for ambulatory crowds where the primary motivation of pedestrians is to travel from a point of origin to a (final or partial) destination [77] [25]. The task of labeling trajectories under a valence-axis scale is made by using the assumption that pedestrians have the motivation to reach a destination, and their emotional valence is influenced by the deviation between expected and actual effort involved in fulfilling this motivation. Within the scope of this research, we refer to expectation as the distance-time expected to meet a motivation, whereas reality corresponds to the actual distance-time required. Pedestrian valence leans towards the positive spectrum when the actual distance-time is equal or less than expected. Conversely, pedestrian valence inclines in the negative spectrum when the actual distance-time surpasses expectation.

3.2.3 Labeling Procedure

Labels for pedestrian emotions are produced on a continuous-valued scale from 0 (negative) to 1 (positive). The expectation is quantified by the DTM measurement explained in section 6.2.4 of chapter 6. Based on the findings of [76], people in ambulatory crowds have the motivation to travel. Motivations can be further divided into commuting, tourism, business, work, and others, with each having a different desired waiting time expressed by his or her walking speed. Under these considerations, emotional annotations are obtained by the following process:

a. The point of origin or destination of a walking trajectory is associated with a POI, where the list of all POI's is provided.

b. Walking trajectories beginning at POI j and arriving at POI k are clustered to the group (j,k).

c. For each trajectory i in the group (j,k), the $dtm^{i,j\to k}$ is computed using the algorithm 1 and adjusted to $\widetilde{dtm}^{i,j\to k}$ with algorithm 2 using the mean arrival time $\overline{\tau}^{j\to k}$ of all trajectories in the group.

d. The expectation for trajectories with origin j and destination k is expressed by the mean DTM, $\overline{dtm}^{j\to k}$, obtained by computing the mean value at every time instance t of the time series $dtm^{i,j\to k}$.

e. For each trajectory i, the deviation between $dtm^{i,j\rightarrow k}$ and $\overline{dtm}^{j\rightarrow k}$ is measured at every time instance t using equation 6.10.

f. The emotional labels $\{e_0^i, e_1^i, ..., e_t^i\}$ corresponding to pedestrian i are produced by applying equation 6.11 to the deviation computed in the previous step, and with e_{exp} selected according to the context of the situation.

Chapter 4

Crowd Simulation Models

A crowd simulation model aims to replicate the behavior of human agents in physical crowds realistically. This task is of particular interest in crowd psychology, where the focus lies in the movement of the agents and crowd. For example, in understanding the group dynamics of people in sports events, concerts, protests, and daily commuting [16] [17]. Collecting data from real world crowds poses several practical and ethical constrains that can be overcome with the use of simulations: (a) reduces the cost and complexity of conducting experiments when a large number of people is required, (b) allows freedom of choice on the environment by fully controlling the virtual infrastructure where the crowd will be observed, (c) facilitates the generation of ground-truth labels, and (d) removes the ethical concerns of recreating situation where pedestrian may be at risk of physical and psychological harm. These benefits come with the caveat that simulations may lack the intrinsic naturalness of an actual crowds, failing to replicate relevant psychological and behavioral aspects, hence challenging the validity of the generated data. In this chapter, we provide a brief introduction of existing models and describe the crowd simulation model developed by us to produce the datasets used in chapters 5, 6, and 8.

4.1 Existing Crowd Simulation Models

A significant number of models have been proposed to address the topic of simulating crowd. The choice of the type of model depends on the purpose of the study and the aspects of the crowd that are relevant. However, as concluded by

previous surveys [145] [45], most crowd simulation models fall in one of three categories: agent-based models, entity-based models, and flow-based models.

4.1.1 Agent-based Models

In this type of model, crowds are formed by a collection of agents, with each agent allowed a degree of autonomy to behave and interact with the simulated environment and other agents as dictated by pre-established rules. Agents in this model are characterized by a decision-making process that involves only the information they are capable of observing from their surroundings. A predominant model in this category is the SFM [65], which balances the agents' need to comply with both social and physical interactions. The SFM is highly suitable in several domains as it has been substantially validated [115] to accurately describe natural human behavior while preserving a degree of unpredictability.

4.1.2 Entity-based Models

Models of this kind aim to implement a set of rules that will shape the behavior of individual agents to replicate a particular social or psychological phenomenon [145]. In this sense, the agents member of the crowd lack autonomy to decide how to behave. This approach is useful in the study of crowd dynamics patterns like queuing, flocking and jamming.

4.1.3 Flow-based Models

Models in this group focus on replicating the movement of a crowd as a whole [145], therefore the agents in this simulation neglect information about their surroundings when making decisions about their behavior, resulting in a limited variation in the behavior of all agents. Flow-based models are useful in studying pedestrian flows of highly dense crowds, for example, in public safety planning and crowd management.

4.2 Proposed Crowd Simulation Model

Under the scope of this research, we are interested in the behavior of crowds at the individual and collective level, understanding the autonomous behavior of pedestrians and the emergence of crowd dynamics. In this sense, we choose

to base our simulation tool on the SFM [65] and incorporate psychological aspects that are relevant for us. The behavior of crowds is generated utilizing an agent-based SFM, as proposed in [65]. Under this model, two types of forces govern the motion of an agent: desired direction force and interaction force. Additionally, the movement of pedestrians is influenced by the infrastructure of the environment, described by walls and a series of points of interest which serve to indicate origin or destination points.

4.2.1 Points of Interest

A POI represents a physical space in the environment and is defined as

$$P_i = \{p_a, p_b\} \tag{4.1}$$

where P_i is the area of the POI i denoted by a bounding box with points p_a and p_b , both in $\mathbb{R}^2$. Some POIs are used to represent entry or exit areas, whereas others can indicate intermediate detours.

4.2.2 Desired Direction Force

This force serves to indicate the motivation of a pedestrian α to reach the destination point $\vec{r}_\alpha^0 \in P_{dest}$ following the shortest possible path. A path has the shape of a polygon with edges $\vec{r}_\alpha^1, ..., \vec{r}_\alpha^n := \vec{r}_\alpha^0$. Given the current position $\vec{r}_\alpha(t)$ and next edge $\vec{r}_\alpha^k$ of the pedestrian α, the desired direction is defined as

$$\vec{e}_\alpha(t) := \frac{\vec{r}_\alpha^k - \vec{r}_\alpha(t)}{||\vec{r}_\alpha^k - \vec{r}_\alpha(t)||} \tag{4.2}$$

Without any obstacles, the pedestrian moves in the direction $\vec{e}_\alpha(t)$ with the desired speed v_α^0. Deviating from the desired velocity $\vec{v}_\alpha^0 := v_\alpha^0 \vec{e}_\alpha(t)$ to avoid obstacles lead to a tendency to regain the desired velocity within a relaxation time τ_α, hence the acceleration term is given by

$$\vec{F}_\alpha^0 := \frac{m_\alpha v_\alpha^0 \vec{e}_\alpha - \vec{v}_\alpha}{\tau_\alpha} + w_\alpha \tag{4.3}$$

where m_α is the mass of the pedestrian and w_α is Gaussian white noise with an arbitrarily selected signal-to-noise ratio (SNR) to account for variations in the pedestrian's perception of the social forces.

4.2.3 Interaction Force

This is a repulsive force and helps account for the influence of other pedestrians β over the movement of pedestrian α and the desired to maintain a certain personal space. The repulsive force is represented by vectorial quantities

$$\vec{F}_{\alpha\beta}(\vec{r}_{\alpha\beta}) := -\nabla_{\vec{r}_{\alpha\beta}} V_{\alpha\beta}[b(\vec{r}_{\alpha\beta})] \tag{4.4}$$

assuming the repulsive potential $V_{\alpha\beta}(b)$ to be a monotonic decreasing function of b with equipotential lines in an ellipse shape, which accounts to the space required for the next step, hence b describes the axis of the ellipse by

$$2b := \sqrt{(||\vec{r}_{\alpha\beta}|| + ||\vec{r}_{\alpha\beta} - v_\beta \Delta t \vec{e}_\beta||)^2 - (v_\beta \Delta t)^2} \tag{4.5}$$

given that $\vec{r}_{\alpha\beta} := \vec{r}_\alpha - \vec{r}_\beta$ and $s_\beta := v_\beta \Delta t$ is of the order of step width of pedestrian β. Additionally, a distance is maintained from walls and other infrastructure obstacles to avoid a collision or getting hurt. Hence, obstacle B results in a repulsive force defined in the form

$$\vec{F}_{\alpha B}(\vec{r}_{\alpha B}) := -\nabla_{\vec{r}_{\alpha B}} U_{\alpha B}(||\vec{r}_{\alpha B}||) \tag{4.6}$$

with a repulsive and monotonic decreasing potential $U_{\alpha B}(||\vec{r}_{\alpha B}||)$ and a vector $\vec{r}_{\alpha B} := \vec{r}_\alpha - \vec{r}_B^{\,\alpha}$ where $\vec{r}_B^{\,\alpha}$ indicates the location of the infrastructure piece B closest to pedestrian α. The total force exerted on pedestrian α is given by

$$\begin{aligned} \vec{F}_\alpha(t) := \vec{F}_\alpha^0(\vec{v}_\alpha, v_\alpha^0 \vec{e}_\alpha) + \sum_\beta \vec{F}_{\alpha\beta}(\vec{e}_\alpha, \vec{r}_\alpha - \vec{r}_\beta) + \\ \sum_B \vec{F}_{\alpha B}(\vec{e}_\alpha, \vec{r}_\alpha - \vec{r}_B^{\,\alpha}) \end{aligned} \tag{4.7}$$

4.2.4 The Pedestrian

While observable, the pedestrian α has a position x_α in $\mathbb{R}^2$, entering the environment at $x_{\alpha_0} \in P_{entry}$ aiming to reach a destination point $x_{\alpha_k} \in P_{dest}$ within an expected time, for which it chooses a particular walking speed v_α^0. Its change in position is denoted by

$$x_{\alpha_t} = x_{\alpha_{t-1}} + v_{\alpha_t} \Delta t \tag{4.8}$$

and the instantaneous walking speed v_{α_t} is given by

Symbol	Name	Description
Φ_1	Ambulatory	A pedestrian walks towards a destination POI while maintaining a personal space and allowing for directional deviation to avoid collisions.
Φ_2	Queue	A pedestrian acknowledges and obeys a queueing structure to be maintained until the destination POI is reached.
Φ_3	Spectator	A pedestrian approaches to a POI and remains in a spectator area for a random period of time.
Φ_4	Panic	A pedestrian approaches to a POI in a fast pace, neglecting the personal space of other pedestrians and potential collisions.

Table 4.1: List of walking behavior types employed by pedestrians to reach a POI.

$$v_{\alpha_t} = v_{\alpha_{t-1}} + \frac{\vec{F}_\alpha(t)\Delta t}{m_\alpha} \tag{4.9}$$

where m_α is the mass of pedestrian α, and $\vec{F}_\alpha(t)$ is as defined in equation 4.7. The motivation of pedestrian α at time t is defined as

$$M_t^\alpha = \{P_{dest}, \Delta T, \Phi_k\} \tag{4.10}$$

where P_{dest} is a POI previously identified, ΔT indicates the maximum amount of time a pedestrian is willing to pursue a motivation before changing its mind, and Φ_k indicates the type of behavior to be employed in reaching P_{dest}. The implemented types of behavior are listed in table 4.1. A pedestrian must have one or more motivations, described by the vector

$$M^\alpha = \{M_1^\alpha, M_2^\alpha, ..., M_n^\alpha\} \tag{4.11}$$

where n is the total number of motivations. A pedestrian exerts the desired direction force in the direction of the current motivation and will continue in this direction until its destination is reached or the maximum time ΔT to meet this motivation expires, at which point the pedestrian pursues the next motivation or is removed from the environment if M^α is empty. The pedestrian parameters considered in this simulation model are summarized in table 4.2.

Parameter	Description
v^0	The desired walking speed to maintain with momentary deviation to avoid collisions.
τ	Relaxation time to regain the desired walking speed.
m	The mass of a pedestrian, used for instantaneous walking speed.
b	Ellipse's axis of personal space to compute repulsive.
P_{entry}	Point of entrance in the environment
M_t	Current motivation of a pedestrian.
P_{dest}	Desired destination associated to the current motivation.
Δ_T	Maximum time allowed to reach the desired destination.
Φ_k	The type of behavior to employ to reach the desired destination.

Table 4.2: List of pedestrian parameters used in the simulations.

4.3 Conclusions

The models presented in section 4.1 prescribe how crowd behavior emerges by limiting the autonomy of pedestrians to different degrees of freedom. An agent-based model sets independent parameters for each agent, relinquishing direct control on what crowd behavior would emerge. Flow-based models establish rules to govern the global behavior of the crowd, greatly limiting the autonomy of the agents member of the crowd. A single framework integrating agent-based and flow-based aspect results in direct conflict. However, a middle ground can be found in Entity-based models, where global rules are imposed on agents while still allowing some degree of freedom. The model introduced in section 4.2 is mainly an agent-based model with a minor entity-based control, handling the pedestrian's behavior by independent parameters shown in table 4.2, with the exception of parameter Φ_k which guides collective behavior on a subset of pedestrians.

Part II

PEDESTRIAN MODEL

Chapter 5

A Pedestrian Model for Emotion Estimation in Crowds

5.1 Introduction

As mentioned in the introductory chapter, the concepts of motivation, expectation, and emotion are the building blocks of this research. The motivation relates to the pedestrian's intended destination, the consequent expectation is an attempt to anticipate the future walking trajectory, and the emotion is evoked from the dissimilitude between expected and actual walking trajectory. These building blocks are present in this and all subsequent chapters, albeit in different forms. In section 5.2 of this chapter, we introduce the first iteration of the pedestrian model for emotion inference based on walking behaviors, where we consider the generic scenario in which pedestrians aim to travel from the point of origin to a final destination. We assume that we can observe and follow the trajectory of pedestrians present in the environment using a surveillance camera or other available sensor capable of detecting people and track their position within an acceptable degree of confidence. We address three main goals in this first iteration of the pedestrian model: (a) provide a representation of the environment, (b) build behavior models based on an environment representation capable of capturing the motivation and expectation of pedestrians, and (c) establish the association between motivations, expectations, and emotions. We integrate these goals by the HDBN presented in figure 5.1.

A representation of the environment serves as a mean to describe locomo-

tive behaviors as a transition between spatial zones in the environment. We employ a self-organizing map (SOM) for this purpose as this method is capable of learning in a data-driven and unsupervised way the existing space for pedestrians to travel and divide it into zones. Furthermore, it provides the means to classify individual observations of pedestrians and associate them to a discrete and mutually-exclusive zone. The obtained topological representation, along with the observed pedestrian trajectories, are used to build a HDBN to model several aspects of pedestrian behaviors. Multiple levels of abstraction are defined in this network for the ease in explaining the semantic and dynamics of each component. At the two lower layers, we define continuous-valued observed and actual states for the pedestrian position in space. At a discrete-valued super-state level where super-states correspond to spatial zones provided by the topological representation, the spatial location of pedestrians are represented in a less granular form, allowing to learn similarity between pedestrian trajectories. In the next level, we define sequences of super-states and grouped into words that correspond to particular walking patterns. In the final level, we assemble sets of words with similar origin-destination into vocabularies, where each specific destination signifies the pedestrian's motivation. This construction of a HDBN allows us to infer relevant aspects of a pedestrian's behavior, such as its short-term movement and subsequent path capturing expectations, and intended destination reflecting his/her motivation. Given the identification of motivation and estimation of expectation, we proceed to infer the pedestrian emotion.

In the remaining content of this chapter, we provide a detailed explanation of the proposed pedestrian model, an experiment to assess the effectiveness of the model, to finally present our conclusions.

5.2 Method

This section proposes a pedestrian model based on a HDBN which can describe the behaviors and associated emotional states of pedestrians. In this work, we define behavior as to how pedestrians transit among different states to achieve its motivation designated by a final destination. For a pedestrian, a state corresponds to its location (x-y coordinates) in a physical region of the environment. Behaviors of pedestrians are labeled empirically by a human operator knowledgeable of the environment using the labels of positive, normal, or negative to denote the emotional state, plus an abnormal state when no label is determined. In overall, our approach starts by learning the topology of the observed envi-

ronment from the trajectory of pedestrians using a SOM [72], which divides the physical space into zones. We represent trajectories of pedestrians as transitions of zones, and all trajectories with a similar destination are classified to the same behavior, to build a probabilistic model that describes this behavior finally. The HDBN for the pedestrian model is presented in figure 5.1. Once the topology of the environment and behaviors of the pedestrians are learned, we can test the ability of the model to produce an estimation of emotions.

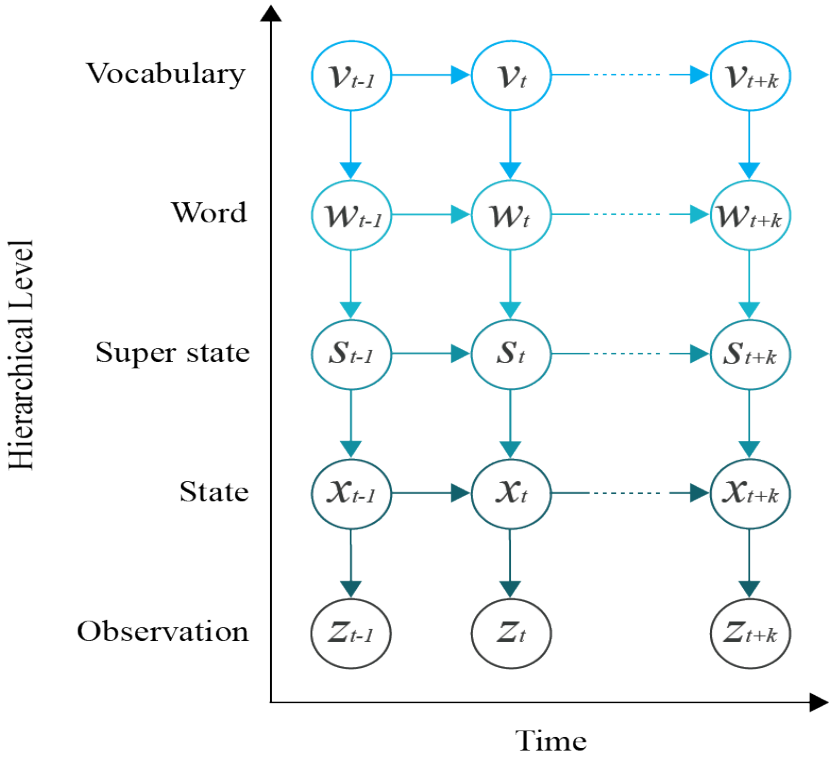


Figure 5.1: HDBN of the pedestrian model.

5.2.1 Environment Representation

We first address the problem of obtaining a topological representation of the environment of interest as this is necessary to describe behaviors of pedestrians. Let us consider an environment monitored by a surveillance camera that captures the motion of pedestrians, as illustrated in figure 5.2a. By applying state of the art techniques for multi-target tracking in camera networks [3] [4] it is possible to obtain the trajectory of each pedestrian as shown in figure 5.2b, and collect this data into a training set X_{train}

$$X_{train} = \{X_1, X_2, ..., X_n\} \tag{5.1}$$

where $X_i = \{x_t, x_{t+1}, ..., x_{t+k}\}$ is the trajectory of pedestrian i over an observation period from t to $t+k$, for a total of n pedestrians. Using X_{train} we train a self-organizing map SOM_p composed by the following elements

- t is the index of target input data vector $x_t \in X_i$ in the training set X_{train}.
- $x_t \in X_i$ is the target input data vector in the training set X_{train}.
- $S = \{s_1, s_2, ..., s_n\}$ is the set of neurons in SOM_p.
- $V = \{v_1, v_2, ..., v_n\}$ is the set of parametric vectors where v_k maps to neuron s_k.
- k is the index of best-matching unit (BMU) in SOM_p.

The set of nodes $S = \{s_1, s_2, ..., s_n\}$ is arranged in an hexagonal topology. Having an input data space $\mathbb{R}^n$, a parametric vector $v_k \in \mathbb{R}^n$ is learned to group all similar input vectors X_t and map them to a node s_k by finding the BMU, designated to be the node with the minimal Euclidean distance

$$\begin{aligned} ||X - v_k|| &= \min_i \{||X - v_i||\} \\ s_k &= \underset{i}{\operatorname{argmin}} \{||X - v_i||\} \end{aligned} \tag{5.2}$$

rewritten for simplicity as

$$s_i = SOM_c(X) \tag{5.3}$$

As a result, the SOM provides a complete topological representation of the environment, where node $s_k \in S$ represents a mutually exclusive zone in the environment, as shown in figures 7.2a and 7.2b. Representing the environment utilizing a SOM encompasses several advantages including (a) unsupervised learning of the environment's topological configuration, (b) clustering and reduction of data, and (c) a simpler way to describe pedestrian trajectories.

5.2.2 Pedestrian Behavior

A walking behavior exhibited by a pedestrian constitutes a response to particular intentions as well as interactions with the environment. To capture this in the context of the present work, we describe each observed pedestrian i in the environment with an instance of the HDBN hence in a crowd with n pedestrians detected we implement a total of n instances. The hierarchical model of the pedestrian is presented in figure 5.1. Starting at the two lower levels of the HDBN, we describe the trajectory X_i of pedestrian i as a discrete-controlled process with continuous-valued state vector $x_t \in \mathbb{R}^2$

$$x_t = F_t x_{t-1} + B_t u_t + g_t \tag{5.4}$$

and observation vector $z_t \in \mathbb{R}^2$

$$z_t = H_t x_t + h_t \tag{5.5}$$

Where F_t is the state transition model applied to the previous state, B_t is the control-input model applied to the control vector u_t, H_t is the observation model, g_t and h_t represent the process and observation noise, both assumed to be independent, Gaussian white. Applying an extended Kalman filter (EKF) over the observation and state vectors, we obtain an estimation $\hat{x}_t$. Details on the EKF are presented in section 6.2.2 of chapter 6. The trajectory X_i of pedestrian i is described as a sequence of estimations

$$X_i = \{\hat{x}_t, \hat{x}_{t+1}, ..., \hat{x}_{t+k}\} \tag{5.6}$$

capturing the short-term movement of pedestrians within a zone of the environment. Proceeding to the next level in the hierarchy, using the zones of S and SOM_p produced in subsection 5.2.1, we can cluster every estimation $\hat{x}_t \in X_i$ into a zone $s_t \in S$ as $s_t = SOM_p(\hat{x}_t)$. This level of abstraction enables us to express the trajectory X_i as a sequence of transitions among zones

$$w_i = \{s_t, s_{t+1}, ..., s_{t+k}\} \tag{5.7}$$

where w_i is called a word. Using words to represent trajectories simplifies the task of finding and clustering similar trajectories and will be of use in the next subsection to define motivations.

5.2.3 Pedestrian Motivation

In the context of this chapter, the motivation of a pedestrian is conceptualized as the intent to reach a physical zone $s_\beta \in S$ in the environment. In the top hierarchical level, words are grouped into a vocabulary k given the condition that $\forall w_a, w_b \in V_k,\ s_1 \in w_a = s_1 \in w_b$ and $s_n \in w_a = s_m \in w_b$ where $|w_a| = n$ and $|w_b| = m$, that is, words with similar origin and destination. The notion of words provides a simplified way to describe trajectories whereas the use of vocabularies allows to group trajectories that correspond to the same origin-destination. Using the subset of words assigned to a vocabulary k, we can proceed to learn a transition matrix

$$V_k = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} \end{pmatrix} \tag{5.8}$$

where $a_{i,j}$ represents the probability of transitioning from zone i to j. From observing a partial word $\tilde{w} = \{s_1, s_2, ..., s_i, s_j\}$ corresponding to the trajectory of a pedestrian and by applying Bayes rule we can predict the zone where the pedestrian is most likely to move next

$$\begin{aligned} a_{i,j} &= P_k(s_i|s_j) \\ &= \frac{P_k(s_j|s_i)P_k(s_i)}{P_k(s_j)} \end{aligned} \tag{5.9}$$

where P_k is a probability mass function (PMF) based on V_k. The probability of observing $\tilde{w}$ under P_k is computed by

$$
\begin{aligned}
P(k, \tilde{w}) &= P(k|\tilde{w}) \\
&= P_k(s_1, s_2, ..., s_j) \\
&= P_k(s_j|s_1, s_2, ..., s_{j-1}) P_k(s_1, s_2, ..., s_{j-1}) \\
&= P_k(s_j|s_1, s_2, ..., s_{j-1}) P_k(s_{j-1}|s_1, s_2, ..., s_{j-2}) P_k(s_1, s_2, ..., s_{j-2}) \\
&= ... \\
&= \prod_j P_k(s_j|s_{j-1})
\end{aligned}
\tag{5.10}
$$

It follows that the motivation of pedestrian i at time t is $m_t = k$ if V_k and P_k best explain the partially observed word $\tilde{w}$ as determined by the argument of the maxima.

5.2.4 Pedestrian Expectation

Along with the inclination of a pedestrian to achieve a motivation conceptualized as the arrival to a physical area in the environment, the expectation serves as the assessment of the effort necessary to accomplish such motivation. In this iteration of the pedestrian model, we define expectation as the time Δ_t required for pedestrian i to fulfill its present motivation m_t. In here, for the sake of simplicity, we work under the rigid assumption that all pedestrians starting at a particular zone s_α with the intention to reach a destination s_β have a similar assessment of the time Δ_t required to reach s_β. Using the subset of words that correspond to a particular vocabulary k we can learn T_k

$$
T_k = \begin{pmatrix}
\mu_{1,1} & \mu_{1,2} & \cdots & \mu_{1,n} \\
\mu_{2,1} & \mu_{2,2} & \cdots & \mu_{2,n} \\
\vdots & \vdots & \ddots & \vdots \\
\mu_{n,1} & \mu_{n,2} & \cdots & \mu_{n,n}
\end{pmatrix}
\tag{5.11}
$$

where $\mu_{i,j}$ represents the mean transition time pedestrians require to go from zone i to zone j when their origin-destination corresponds to s_α and s_β as describe by vocabulary k. The expectation Δ_t for the trajectory of pedestrian i expressed as a word $w_i = \{s_t, s_{t+1}, ..., s_{t+n}\}$ is computed as

$$
\Delta_t = \sum_{t=0}^{n-1} \mu_{t,t+1}
\tag{5.12}
$$

where $\mu_{t,t+1}$ is obtained from T_k to obtain the mean time required to travel from s_t to s_{t+1} for all zones in w_i. Similarly, when we are presented with a partially observed word $\tilde{w}_i$, we make use of equation 5.9 and maximum likelihood to estimate the word $\hat{w}_i$ that better predicts the whole trajectory to be taken by pedestrian i, and on the sequence of transitions denoted by $\hat{w}_i$, we proceed to compute its expectation $\hat{\Delta}_t$.

5.2.5 Pedestrian Emotion

At this point in our research, the emotion E_i of a pedestrian i is measured by one of three possible discrete states (positive, negative, and neutral) using the Positive/Negative scheme introduced in section 3.1.3 of chapter 3. Having empirical labels of emotional states on all trajectories of the training set corresponding to the above mentioned discrete states, as provided by a human expert, we proceed to learn three models as described in equation 5.11:

- T_k^+ is learned using the trajectories of vocabulary k labeled as positive.
- T_k^* is learned using the trajectories of vocabulary k labeled as neutral.
- T_k^- is learned using the trajectories of vocabulary k labeled as negative.

Given a partially observed word $\tilde{w}$ and its corresponding elapsed time τ, the expected times Δ_t^+, Δ_t^* and Δ_t^+ derived from each emotional model are computed as prescribed in equation 5.12. The task of emotion inference is addressed as a one-versus-all classification problem

$$E_t = \underset{e \in \{+,*,-\}}{\operatorname{argmin}} \tau - \Delta_t^e \tag{5.13}$$

where the model yielding the smallest difference between expected and elapsed time indicates the estimated emotion E_t of pedestrian i at time t. It is import to note that the association between behavior and emotion depends on the context of the situation; the rules for labeling are to be determined for each particular scenario.

5.3 Experiments and Results

To validate our proposed model we conducted an experiment employing data produced by a realistic crowd simulator first introduced in [28] [29], based on

	Training dataset	Test dataset
Duration (hours)	5	5
Positive trajectories	978	894
Neutral trajectories	1770	1815
Negative trajectories	252	291
Total trajectories	3000	3000

Table 5.1: Details of training and testing datasets produced from simulations.

the social force model [65] where each pedestrian in the environment is treated as a particle subject to forces in a two-dimensional space, deriving its motion equations from Newtons law $F = ma$ and accounting for its motivation as an attraction force pulling the pedestrian towards its destination and repulsive forces from physical objects and other pedestrians in the environment. We have recreated a scenario similar to that of a train station, as shown in figure 5.2a and the produced trajectories are plotted in figure 5.2b. The information of pedestrian trajectories is provided directly from the crowd simulator, hence the steps for pedestrian detection and tracking are omitted. Simulations were carried out under different settings, and details for the training and testing datasets produced from these simulations are presented in table 5.1.

5.3.1 Model Training

The self-organizing map SOM_p is trained with the following configuration: The set of neurons in SOM_p is initialized with random weights and in a hexagonal arrangement spread across the corresponding input space. Distance between neurons is calculated by the number of links among them. The initial neighborhood size is 3, with 100 steps for the ordering phase. The training phase is done over 500 epochs by competitive layer but without bias, updating the winning neuron and all other neurons within the given neighborhood using Kohonen rule. The first addressed task is to use the trajectory of pedestrians to obtain a topological representation of the environment with the help of a self-organizing map SOM_p as shown in figures 7.2a and 7.2b. The distribution of training data among zones after the clustering process is reflected in the learned topology where larger zones indicate that more trajectories traverse this area and the opposite is also true for smaller zones. The decision of how many zones to employ to describe trajectories has a direct impact on the reliability of our model to estimate the emotion of pedestrians; this happens because we represent the be-

havior of pedestrians by the transition of zones, and with fewer zones, there is a higher uncertainty of the motion of pedestrians. In these experiments, SOM_p is composed of 100 zones (10 rows and 10 columns) in a hexagonal topology. After testing our model with different dimensions, we found this size to be a suitable balance between predictability and topological representativeness. Employing SOM_p, the trajectories of the training set were evaluated, and a total of 41 different behaviors were identified, a few examples of the learned behaviors are shown in Figure 5.4.

The scenario replicated in the experiments corresponds to that of a train station. Hence, the criteria for labeling behaviors follows from the assumption that people aim to reach their destination in the briefest possible time. The behaviors with the minimum number of state transitions and the shortest transition time are labeled to correspond to a positive emotion. The behaviors with a higher frequency of occurrence are associated with neutral emotion. Finally, the behaviors with the highest number of transitions and longer transition time are assigned a negative emotion.

5.3.2 Model Evaluation

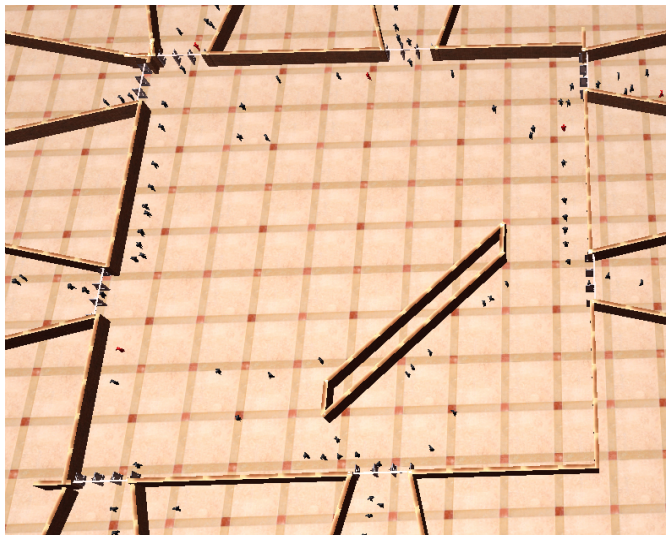
During the testing phase, to estimate the emotional state of the pedestrians, our hierarchical model predicts the zones transitions and transition time for each pedestrian in real time based on the learned behaviors. In our model, the accuracy level to estimate the emotion of pedestrians depends on the model's capability to predict the pedestrian's behavior. In figure 5.5a, we show the behavior prediction accuracy during 100 seconds, where the mean accuracy was 76%. Throughout the entire length of the simulations, the mean accuracy ranged between 74% and 82%. A summary of the model's performance to estimate emotional states is presented in table 5.2. In figure 5.5b, we offer a snapshot of the online emotion estimation of pedestrians. The observed results on emotion estimation show neutral emotion to be the most accurately identified emotion closely followed the remaining two. The majority of misclassified trajectories occur between positive and neutral emotions as expected due to their similarity in behavior and since most of the generated trajectories correspond to these two groups. Trajectories with negative labels are identified at a similar rate than the other emotions despite the significantly smaller number of available negative trajectories in the training and testing phase.

	Positive	Neutral	Negative
Positive	71%	23%	6%
Neutral	14%	83%	3%
Negative	4%	21%	75%

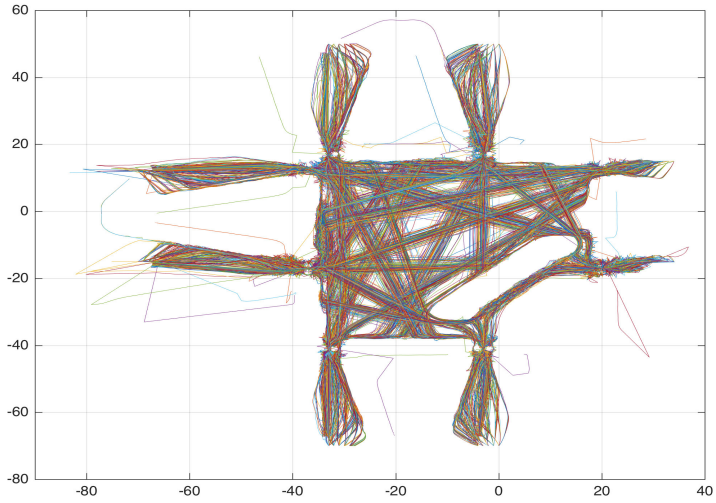
Table 5.2: Confusion matrix of pedestrian emotion estimation based on behavior classification.

5.4 Conclusions

In this chapter, we presented the first iteration of the pedestrian model for the estimation of individual emotions of pedestrians under complex, crowded environments. In comparison with [5], our approach provides significant improvements: (1) accounts for scenarios with multiple origin and destination points, (2) Introduces the concepts of motivations and expectations as the building blocks to estimate emotions. (3) Presents the idea of representing behaviors in higher abstractions using words and vocabularies, which helps to reduce data sparsity. The conducted experiments confirm the viability of this model to estimate pedestrian motivations, model their behavior, and make use of this to produce an estimation of their emotional state. The approach presented here applies to crowded real-life environments for monitoring automation intended to identify and prevent dangerous situations as well as to improve crowd control. Furthermore, contributions of this nature are essential for the development of robust cognitive dynamic systems intended for smart cities. Shortcomings of this method include: (1) the hard assumption that all pedestrian following a particular behavior model has the same expectation and agency, (2) the scenario where pedestrians change their motivation is not accounted for, and (3) the method for labeling trajectories to an emotional state is restricted to only three possible states and is not empirically validated. Future development of this work will focus on extending the model to address the previously mentioned limitations and will consider the interaction of pedestrian and crowd emotions enabling us to explore causality and contagion of emotional states among pedestrians and its impact in the crowd as a whole.

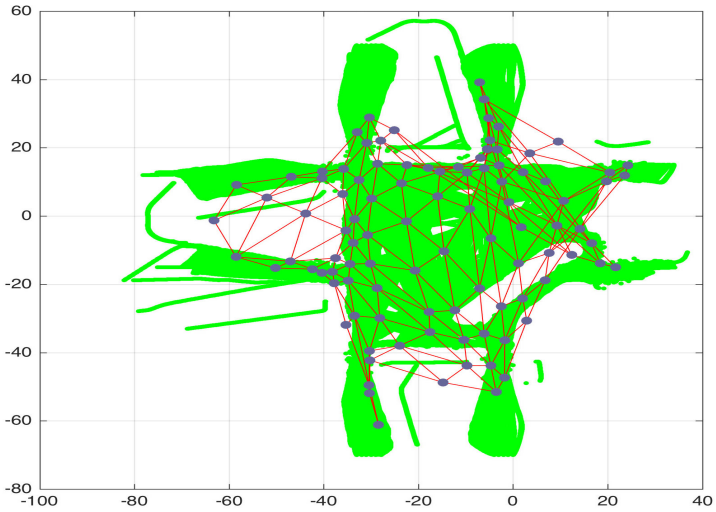


(a)

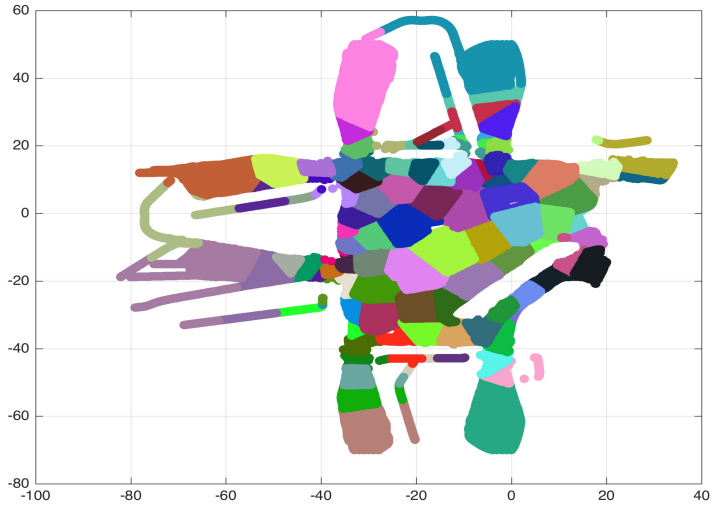


(b)

Figure 5.2: (a) A snapshot of the simulated environment. (b) A plot of pedestrian trajectories with colors assigned randomly.



(a)



(b)

Figure 5.3: (a) Training data (green) and the self-organizing map SOM_p (red edges and blue nodes). (b) Environment partitioned into zones with colors assigned randomly.

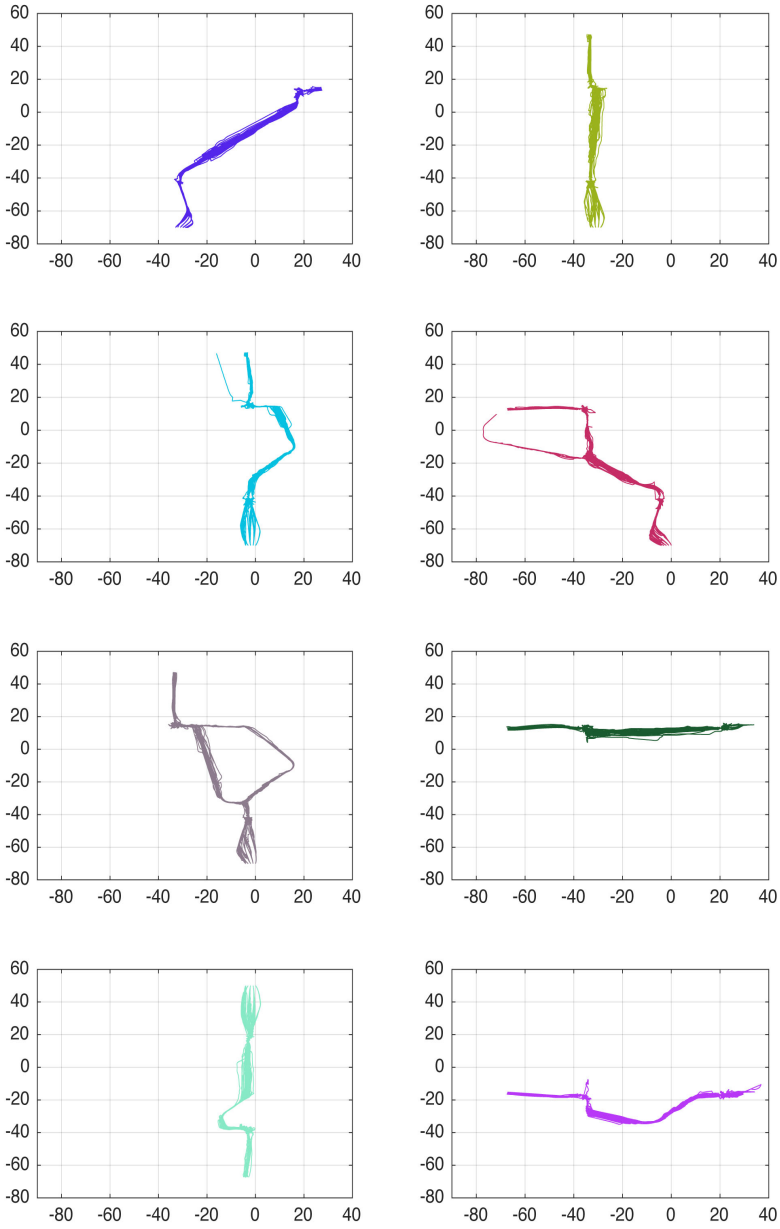
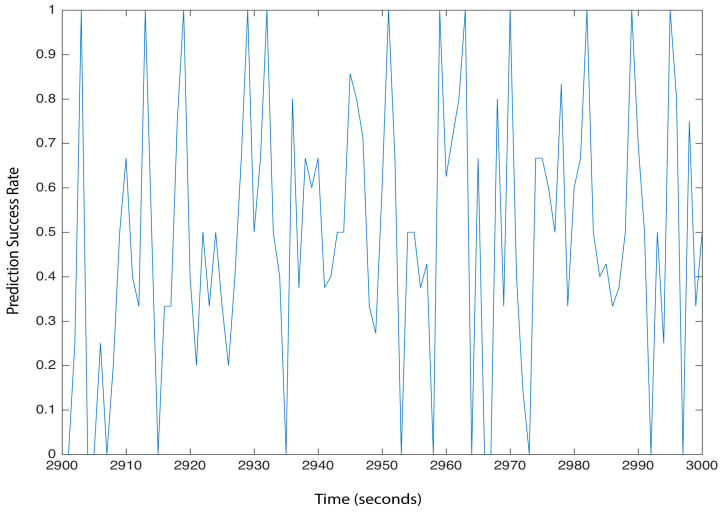
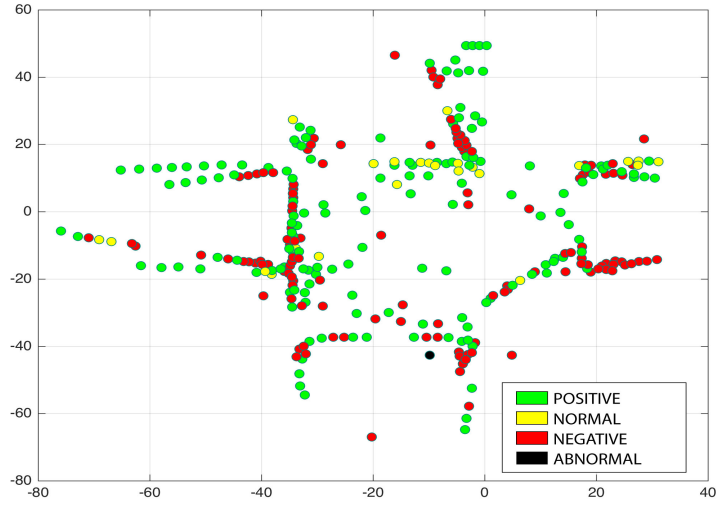


Figure 5.4: Examples of learned pedestrian behaviors from the trajectories in the training phase, a total of 41 different behaviors were identified. Colors are assigned randomly.



(a)



(b)

Figure 5.5: (a) Pedestrian behavior prediction accuracy. (b) Snapshot of online pedestrian emotion estimation.

Chapter 6

Accounting for Motivations and Expectations in the Pedestrian Model

6.1 Introduction

This chapter presents a new iteration of the pedestrian model aiming to enhance the way we capture the interplay of motivations and expectations in inferring pedestrian emotions, as supported by the body of literature in the field of psychology. We start by providing the foundation for the design of the pedestrian model and continue to outline the improvements over the previous iteration of the model. Field theory describes a person's behavior to be compound by different fields, and for each field, there is a motivation, an attracting force guiding the movement towards the desired goal as well as (potential) repulsive forces preventing to reach that goal [20]. In this sense, motivation is the reason for a person to engage in a given behavior and interact with an environment (field) to achieve an objective. Consequently, motivations involve expectations [76], a person's assessment of the effort required to meet motivation. Emotions play an essential role as they regulate a person's behavior in an attempt to make reality match expectations [114]. In the context of a crowded environment, a pedestrian engages in a walking behavior coherent to the intended motivation. The expectation is conceptualized as the desired conditions to fulfill a motivation

and the emotion in the individual arising from the departure between expectation and reality, yielding an emotion in the positive spectrum when reality approaches (or surpasses) expectation or an emotion in the negative spectrum for the opposite case. Incorporating knowledge from force field theory, probability theory, expectation psychology and emotional theories, this iteration of the pedestrian model proposes a data-driven approach capable of inferring the emotion of pedestrians in crowded environments by estimating individual motivations and expectations based on observed walking behaviors. Hence, this method is suited for crowds where pedestrian display ambulatory behaviors. In this chapter, the pedestrian emotion is measured following a Dimensional model approach presented in section 3.2 of chapter 3, in which emotions are expressed as a continuous point in a valence-arousal affect space, however only the valence component is taking into account here. The revised HDBN of the pedestrian model is presented in figure 6.1, and it describes the influence of motivations (as associated with points of interest) on walking behavior of pedestrians, followed by conceptualizing expectations, to finally use these elements to produce an estimation of the emotional state of individual pedestrians.

Significant improvements over the model described in chapter 5 include: (1) The super-state and word levels of the HDBN have been removed, and the latent variable for motivation directly influences the walking trajectory described by the state and observation vectors, (2) a radial direction field is learned for each motivation and used to predict walking behaviors (3) a new measurement to quantify expectation as time and distance traveled is introduced, and (4) the quantification of emotion is extended from discrete to continuous-valued states from positive to negative valence.

The remaining content of this chapter provides a comprehensive description of the pedestrian model along with two experiments aiming to asses the effectiveness of the model on a broader variety of circumstances, and finally, we present our conclusions.

6.2 Method

The method presented here is shown in figure 6.1. At the two lower levels, the observation and state vectors are measured as continuous-valued quantities to represent the short-term movements of the pedestrian. The transition model that aims to predict the future movement of the pedestrian in the state vector is constructed from an EKF where a control unit vector is given by a Radial function, having one separate radial function learned for each POI in

the environment. The right selection of the EKF model depends on the inference of the pedestrian motivation represented at the top level of the HDBN. A conditional probability distribution dictates the transition among motivations. Outside of the HDBN, the component for expectation is computed according to each pedestrian specific parameters, and an estimation of its emotional state is produced at the end of this process. In the training phase, manual annotations of the environment, POI's, and a set of trajectories are used to learn the environment's representation, walking behaviors, motivation and expectation models, and emotion estimation parameters. During the testing phase, partial observations of pedestrian trajectories are feed to the model and inference about their motivation and expectation being computed as necessary to produce an estimation of their individual emotional state. As a precondition preserved from the previous iteration of the models, the pedestrian is observed using a pedestrian detection/tracking technique capable of extracting a significant portion of the pedestrian's trajectory. If this condition fails to hold, the performance of the pedestrian model is considerably affected.

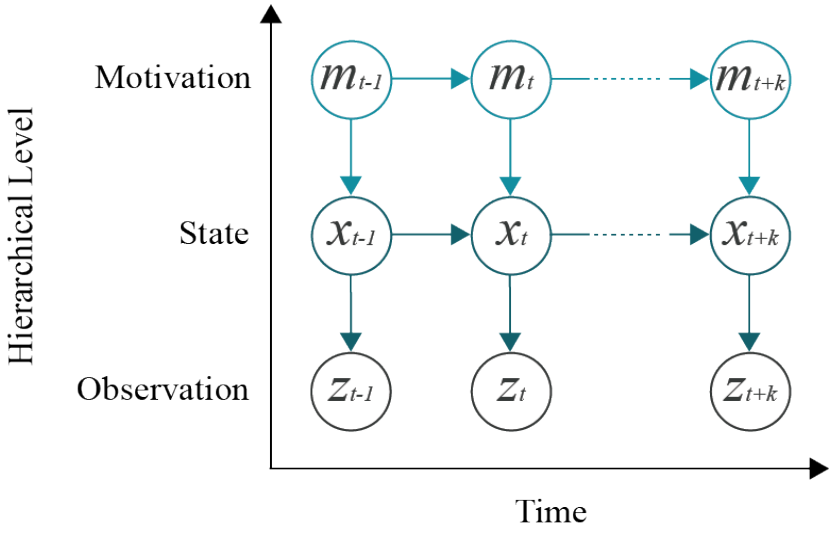


Figure 6.1: HDBN of pedestrian model.

6.2.1 Environment Representation

The scope of this model considers environments intended for crowds where people predominantly display ambulatory behavior. This entails that the main objective of people in such surroundings is to travel from one place to another where each one of these places is called a POI. Formally described, a POI refers to the physical location of an entity that one or many people find useful or interesting. Common examples of POI's in a train station, for instance, include entrance/exit doors, information kiosks, ticketing windows, ATMs and lavatories. A POI i is defined by the following elements:

- $B_i \in \mathbb{R}^2$ is a set of points delimiting a bounding box for POI i.
- $\vec{D}_i(x) = u_i$ is a radial direction field defined as a vector point-function which takes a point $x \in \mathbb{R}^2$ in the space of the environment and returns a unit vector u_i in the direction of POI i

An illustration of the environment representation is shown in figure 6.2 where sub-figure6.2a indicates the labels of the identified POI's, sub-figure 6.2b displays the manually annotated bounding boxes, and sub-figure 6.2c presents an example of a direction field for a particular POI.

6.2.2 Pedestrian Behavior

The state of pedestrian i at time t is described by its position, indicated with the continuous-valued vector $x_t^i = [x, y]^T$ in $\in \mathbb{R}^2$. Its behavior is modeled as an EKF, assumed to be influenced by a motivation $m_t^i = \omega$ where ω is the index of the destination POI. Hence m_t^i is treated as a switching variable, with a different behavior model for each destination

$$x_t^i = F_t x_{t-1}^i + \hat{v}_0^i \vec{D}_\omega(x_{t-1}^i) + u_k \tag{6.1}$$

where F_t is the state transition matrix defined as a unit matrix, $\hat{v}_0^i$ is a scalar indicating the estimated desired walking speed, $\vec{D}_\omega$ the direction field of POI ω, and u_t accounts for the interaction forces exerted on pedestrian i at time t modeled as Gaussian white noise with covariance $Q_t : u_t \sim \mathcal{N}(0, Q_t)$. The observation model is given by

$$z_t^i = H_t x_t^i + v_t \tag{6.2}$$

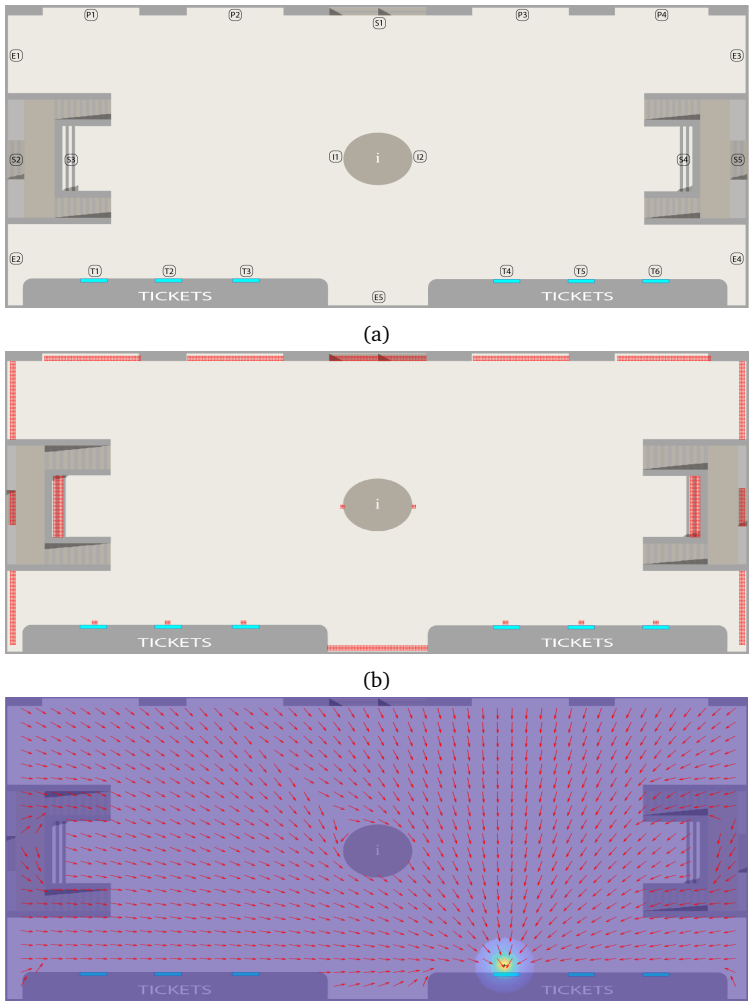


Figure 6.2: Illustration of Grand Central station in New York city displaying (a) the labels of all POI's in the environment, (b) bounding boxes for each POI marked with red dots, and (c) an example of the direction field for a particular POI.

where H_t is the observation model defined as a unit matrix, and v_t corresponds to the observation noise assumed to be zero-mean Gaussian white noise with covariance $R_t : v_t \sim \mathcal{N}(0, R_t)$. Based on the study presented in [76], this model assumes that a pedestrian tends to maintain the desired walking speed $\hat{v}_0^i$, temporarily affected by its interaction with the environment and other pedestrians. Hence, the desired walking speed is taken from the mean walking speed from previous observations. The innovation with respect to the behavior model ω is defined as

$$\tilde{y}_t^{i,\omega} = z_t^i - H_t \hat{x}_{t|t-1}^{i,\omega} \tag{6.3}$$

where $\hat{x}_{t|t-1}^{i,\omega}$ corresponds to the estimation produced using $\vec{D}_\omega$.

6.2.3 Pedestrian Motivation

The motivation $m_t^i = \omega$ describes the intention of pedestrian i at time t to reach the POI ω and is conceptualized as a discrete latent variable inferred from its walking behavior. Therefore is modeled as a hidden Markov model (HMM) with N number of states corresponding to the total number of POI's and transition probability

$$P(m_t^i, \Delta t | m_{t-1}^i) = P(m_t^i | \Delta t, m_{t-1}^i) P(\Delta t | m_{t-1}^i) P(m_{t-1}^i) \tag{6.4}$$

where the transition probability is conditioned on the previous motivation m_{t-1}^i and the elapsed time Δt until the next transition. The previous motivation $m_{t-1}^i = \alpha$ is interpreted as the point of origin and $m_t^i = \omega$ as the intended destination. This work proposes POI ω to function as an attractor for pedestrian i when $m_t^i = \omega$, hence the behavior model using the direction field $\vec{D}_\omega$ is expected to best describe the observed walking behavior. The motivation of pedestrian i is estimated as

$$m_t^i = \underset{\omega \in \{1,\dots,n\}}{\operatorname{argmin}} \sum_{t=a}^{b} \tilde{y}_t^{i,\omega} \tag{6.5}$$

where a partial observation $z_{a:b}^i$ is used to compute the prediction $\hat{x}_{a:b}^i$ and innovation $\tilde{y}_{a:b}^{i,\omega}$ for each available behavior model ω, and the model yielding the smallest innovation is selected as the current motivation.

6.2.4 Pedestrian Expectation

Research in the field of Expectation Psychology [76] points to the idea that a person engages in a particular behavior aiming to fulfill a goal by putting an amount of effort close to its psychological expectation. Expectations are formulated according to a person's previous experience, motivation, and the context of the environment [70]. This research focuses on ambulatory crowds; hence, the purpose of pedestrians is predominantly to travel. In this sense, a pedestrian i starts at a POI α with the motivation to reach the POI ω (i.e. $m_t^i = \omega$), for which from a first-person perspective there is a distance to travel measured as

$$\mathscr{D}_{act}^{i,\alpha\to\omega} = \sqrt{(x_f^i - x_0^i)^2} \tag{6.6}$$

That is, the Euclidean distance between a pedestrian's initial position x_0^i and $x_f^i \in B_\omega$. The actual time to meet this motivation can be derived as

$$\mathscr{T}_{act}^{i,\alpha\to\omega} = \frac{\mathscr{D}_{act}^{i,\alpha\to\omega}}{\bar{v}^i} \tag{6.7}$$

Where $\bar{v}^i$ is the mean walking speed throughout the journey. However, from a pedestrian's perspective, the expected distance is hypothesized (based on observation or previous knowledge) as

$$\mathscr{D}_{exp}^{i,\alpha\to\omega} = \mathscr{D}_{act}^{i,\alpha\to\omega} + r^i \tag{6.8}$$

Accounting for a perception error, r^i assumed to be zero-mean Gaussian white noise. Consequently, the expected time to reach ω is derived from

$$\mathscr{T}_{exp}^{i,\alpha\to\omega} = \frac{\mathscr{D}_{exp}^{i,\alpha\to\omega}}{\hat{v}_0^i} \tag{6.9}$$

with a desired walking speed $\hat{v}_0^i$ chosen according to the level of urgency to meet that motivation and the amount of effort willing to use.

To depict how a pedestrian i with a path x^i approaches towards its desired motivation over time, a DTM $dtm^{i,\alpha\to\omega}$ time series is calculated as described in Algorithm 1. To learn the expected way in which pedestrians with the same POI origin α and destination ω reach their target, the mean DTM $\overline{dtm}^{\alpha\to\omega}$ of all paths in the group (α,ω) is computed. However, as pedestrians walk at different speed hence producing $dtm^{i,\alpha\to\omega}$ with varying arrival times $\tau_{act}^{i,\alpha\to\omega}$, the

$dtm^{i,\alpha\to\omega}$ of all pedestrians in the group (α,ω) are first adjusted to $\widetilde{dtm}^{i,\alpha\to\omega}$ using Algorithm 2 where $\overline{\tau}^{\alpha\to\omega}$ is taken from the mean arrival time. It follows that $\overline{dtm}^{\alpha\to\omega}$ shows the normality of how a pedestrian approaches its motivation, and it will be used in the next section for estimating the emotional state of a pedestrian.

Algorithm 1 Compute DTM of one trajectory

Input:
1: $[x^i]$ Pedestrian trajectory
2: $[x^i_f]$ Final position
Output:
3: $[dtm^i]$ DTM time series of pedestrian
4: **procedure** COMPUTE DTM
5: **for** $k = 1$ **to** $length(x^i)$ **do**
6: $dtm^i_k \leftarrow \sqrt{(x^i_f - x^i_k)^2}$

6.2.5 Pedestrian Emotion

In the presented model, the emotion of a pedestrian is represented in a single-axis valence with a continuous value ranging from 0 (negative) to 1 (positive). The central idea to estimate the emotional state of a pedestrian is by measuring the deviation between expectation and reality for a particular motivation. Expectation is represented by $\widehat{dtm}^{i,\alpha\to\omega}$ which is computed by taking $\overline{dtm}^{\alpha\to\omega}$ and adjusting it to an expected arrival time $\tau^{i,\alpha\to\omega}_{expected}$ obtained from equation 6.9. Reality is given by the actual $dtm^{i,\alpha\to\omega}$ computed as time passes. The expectation is further reduced to a numerical value AUC^i_{exp} by applying the trapezoidal rule over the curve described by $\widehat{dtm}^{i,\alpha\to\omega}$. In the same way, the actual observed behavior captured in $dtm^{i,\alpha\to\omega}$ is reduced to AUC^i_{act}. The deviation between AUC^i_{exp} and AUC^i_{act} is measured by

$$AUC^i_{\Delta} = \frac{AUC^i_{exp} - AUC^i_{act}}{AUC^i_{exp}} \tag{6.10}$$

and the emotional state e^i_t of pedestrian i at time t is computed as

Algorithm 2 Adjust DTM of one trajectory to the desired arrival time

Input:

[dtm^i] DTM time series of pedestrian
[$\tau^{i,\alpha\to\omega}_{actual}$] Actual arrival time of DTM
[$\overline{\tau}^{\alpha\to\omega}$] Adjusted arrival time of DTM
[δ_t] Size of the time interval

Output:

[$\widetilde{dtm^i}$] Adjusted DTM time series of pedestrian
procedure ADJUST DTM
    $n \leftarrow \tau^{i,\alpha\to\omega}_{actual}/\delta_t$, $m \leftarrow \overline{\tau}^{\alpha\to\omega}\ \delta_t$
    $dtm\_idx \leftarrow round([1:n] * (m/n))$
    $dtm\_idx_1 \leftarrow 1, dtm\_idx_n \leftarrow m$
    for $t = 1$ **to** m **do**
        $h \leftarrow$ first index value where $dtm\_idx_h >= t$
        if $(t == 1)$ **or** $(t == dtm\_idx_h)$ **then**
            $\widetilde{dtm}_t \leftarrow dtm_h$
        else
            $p = (t - dtm\_idx_{h-1})/(dtm\_idx_h - dtm\_idx_{h-1})$
            $\widetilde{dtm}_t = dtm_{h-1} - (dtm_{h-1} - dtm_h) * p$
        end if
    end for
end procedure

$$\begin{aligned} e_t^i &= e_{exp} + e_{exp} f(AUC_\Delta^i) \\ &= e_{exp}(1 + f(AUC_\Delta^i)) \end{aligned} \tag{6.11}$$

where e_{exp} is the expected emotion manually associated to this DTM, and $f(x)$ is a standard logistic regression function with parameters $k = 1$, $x_0 = 0$, and $L = 1$

$$\begin{aligned} f(x) &= \frac{L}{1 + e^{-k(x - x_0)}} \\ &= \frac{1}{1 + e^{-x}} \\ &= \frac{e^x}{e^x + 1} \end{aligned} \tag{6.12}$$

This way of computing pedestrian emotions is constrained by how well the motivation is estimated but is advantageous in the sense that previous wrong estimations of m_t^i don't influence future values of e_t^i because of the expectation values $\widehat{dtm}^{i,\alpha\to\omega}$ and AUC_{exp}^i are recalculated on each new estimation of m_t^i.

6.3 Experiments and Results

To validate the presented model, we proceed to conduct two experiments covering real-world and simulated data in a wide variety of circumstances to better explore the adaptability of the proposed model.

6.3.1 Experiment 1: C-Station Dataset

Published by [137], the C-Station dataset captures footage of the main concourse at Grand Central station in New York City and provides manual annotations of the pedestrians trajectories visible over the period of observation. Details about the date-time of the footage are not provided; however, the observed conditions appear to correspond to non-peak hours with density predominantly

Resolution (pixels)	1,920 x 1,080
Total Frame Number	100,000
Frame Rate (fps)	25
Annotated Frame Number	5,000
Annotated Frame Rate (fps)	1.25
Annotated Pedestrian Number	12,684
Average Pedestrian Number per Frame	123
Maximum Pedestrian Number per Frame	332

Table 6.1: Details of C-Station dataset [137].

low. More information about this dataset is summarized in table 6.1, and figure 6.3 presents the observed environment. In the following evaluation of the pedestrian model, the annotated trajectories of this dataset are separated in 70% for training and 30% for testing.

Model Training

During the training phase, 10 POI's are identified, and a radial direction field $\vec{D}_i(x) = u_i$ is defined for each POI in the environment, and they will serve for the walking behavior models. The training set trajectories are clustered by a pair

Figure 6.3: The observed environment in C-Station dataset with POI's denoted with yellow rectangles and sample pedestrian trajectories plotted in randomly colored lines.

of origin-destination POI's, and DTM's are computed as described in subsection 6.2.4, the process is illustrated in figure 6.4. Finally, the expected emotion $e_{exp} = 0.5$ is used for each DTM.

Model Evaluation

In the testing phase, the pedestrian model is evaluated by its accuracy to predict pedestrian motivations which enable the selection of the appropriate DTM, inference about the desired walking speed, and finally the emotional state estimation. Predicting the pedestrian's motivation consist in determining the destination POI based on partial observations of a trajectory. Similarly to [137], 10 POI's are identified as possible origin/destinations, and the first half of each observed trajectory is taken as input in the experiment. The methods [137] and [144] are used for comparison.

The proposed method is evaluated using different values of θ arbitrarily selected, where θ represents the number of past observations used to estimate the motivation. In the case of $\theta = 0$, the innovation $\tilde{y}^{i,k}$ is accumulated for the whole observation period whereas, for $\theta > 0$, the innovation is accumulated using only the last θ number of observations. Results of predicting pedestrians motivations are presented in figure 6.5. From these results, we can observe the optimal value of $\theta = 5$ where the maximum accuracy of 72% is achieved, significantly outperforming the compared methods [137] and [144]. This outcome evidences that, at least in this dataset, including all observations affects performance as pedestrians do not necessarily choose the most optimal path to reach their destination or may change their motivation. Small values of θ can tamper accuracy if several POIs are in a similar direction. An optimal value for θ allows for enough evidence to improve accuracy while ignoring potentially misleading segments of a pedestrian path.

The estimation of emotion under the proposed hypothesis is subjected to how well the motivation m_t^i and desired walking speed $\hat{v}_0^i$ are estimated. Once the motivation is determined, the desired walking speed is taken from the mean walking speed from previous observations as we assume a tendency on the pedestrian to oscillate towards the desired walking. The MSE measurement is adopted to evaluate the emotional state estimation. The results of emotion estimation are presented in figure 6.6, under different values of parameter θ for motivation prediction. The minimum MSE is achieved when $\theta = 50$, yet this difference is not significant compared to $\theta = 5$, which is found to be the optimal value for motivation prediction. We conclude that once the motivation is

predicted correctly, the emotion estimation consistently improves with an increasing number of observations available.

6.3.2 Experiment 2: Grand Central Station Synthetic Dataset

Aiming to expand the limited diversity of scenarios presented in [137], an agent-based social force model was implemented in this experiment to accurately replicate the observed environment in experiment 1 and to conduct simulations that expose the crowd to a broader variety of circumstances. Details of the crowd simulation model are presented in chapter 4 and the methodology introduced in section 3.2 of chapter 3 is used in this dataset to produce emotional annotations. The produced dataset captures common scenarios such as peak and non-peak hours as well as abnormal ones like panic situations. Details of the SC-Station dataset are summarized in table 6.2, and the simulated environment is presented in figure 6.7. All contexts of the SC-Station dataset are divided into training and test sets for the experiments performed in this section.

Model Training

In the training phase of the pedestrian model, a total of 22 POI's are identified from manual annotations as seen from figure 6.7 and a radial direction field $\vec{D}$ is generated for each POI. Different from experiment 1, here we consider temporal POI's as pedestrians have intermediate destinations before exiting the environment (e.g., ticket windows and information kiosks). Trajectories are grouped by origin-destination POIs to learn DTM's and expected emotions e_{exp} are assigned accordingly to each context, with low valence in the panic scenario ($e_{exp} = 0.1$), below neutral for peak hours ($e_{exp} = 0.4$) and neutral for non-peak hours ($e_{exp} = 0.5$).

Model Evaluation

In the testing phase, the pedestrian model is evaluated over all four contexts of the SC-Station dataset, starting with the prediction of pedestrian motivations to allow the appropriate selection of DTM used to quantify expectation and subsequently the estimation of pedestrian emotion. Pedestrian motivation prediction is made using all past observations of a pedestrian trajectory ($\theta = 0$) as this parameter value yielded the best results in the first experiment. Figure 6.8a presents the mean accuracy in predicting pedestrian motivations for each context. In general, all contexts show a lower accuracy with respect to experiment

1, which is expected as the environment in experiment 1 identifies 10 POI's in contrast to the 22 POI's identitied in experiment 2, leading to a reasonable conclusion that an increase in the number of POI's causes higher similarity between walking behaviors, directly impacting the motivation prediction accuracy.

Other important factors influencing the accuracy in motivation prediction can be attributed to the number of pedestrians present in the environment and their walking speed, this is explained by a pedestrian's need to frequently and rapidly change its walking direction to avoid collision with other pedestrians when density and average walking speed is higher. This insight is confirmed from observing that the non-peak hours' context has the highest accuracy (0.60) in motivation prediction while it also has the lowest mean/max walking speed (0.38/0.52) and the second lowest density (896). Conversely, the evening peak hours context yields the lowest accuracy in motivation prediction (0.31) but counts with the highest number of pedestrians (4559) and the second fastest mean/max walking speed (1.52/2.01).

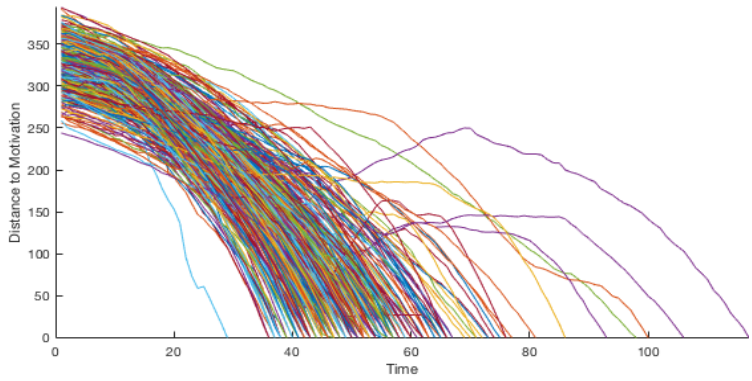
The accuracy to estimate pedestrian emotions is measured by computing the MSE between the labeled emotion e^i and the estimated emotion $\hat{e}^i$ for each trajectory i. The mean MSE of pedestrians in each test set is presented in figure 6.8b.

In general, higher accuracy in pedestrian motivation should yield a lower MSE in estimated emotion if pedestrians exhibit a behavior close to their estimated expectation; however, this was not the case for three out of four test sets as seen from figure 6.8. A possible interpretation of this outcome is that the panic scenario holds an inverse correlation between motivation prediction accuracy and emotion estimation MSE because, under this scenario, pedestrians unsurprisingly rush to the closest exit in a more predictable behavior. In the case of the evening peak hours context it is essential to consider that whereas the motivation prediction accuracy may be lower, the selected DTM to compute expectation may not be so different if the distance between the actual and predicted motivations is small, hence still producing an emotion estimation close to its labeled emotion.

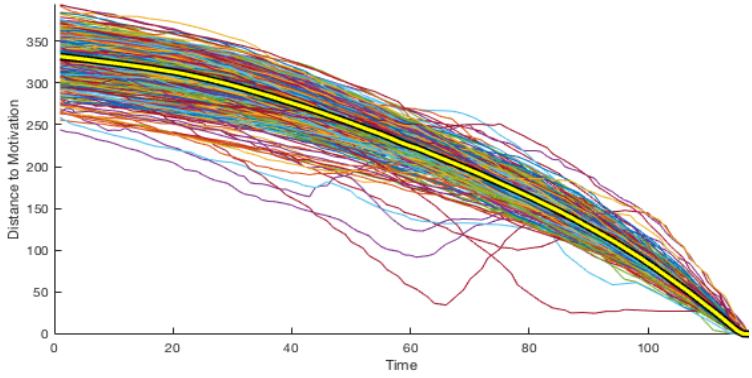
In summary, the number of POIs in the environment is an important factor in motivation prediction, whereas density and walking speed greatly influence the estimated expectation and consequently, the inference of pedestrian emotions.

6.4 Conclusions

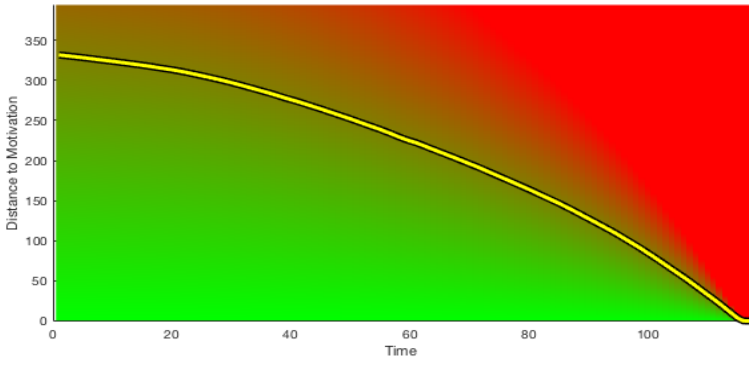
This chapter introduced an improved version of the pedestrian model intended for estimating individual emotions based on observed walking behavior and inferred motivations and expectations. The prediction of pedestrian's motivation was addressed using direction fields generated for each POI. The emotional state is derived from the difference between expected and actual observed behaviors. A hypothesis of expectation for pedestrians in a train station was proposed and employed to generate emotional state annotations. The results of the proposed model indicate a significant improvement for predicting motivations (destinations) over previous works, and to efficiently estimate individual emotions based on the proposed hypothesis for expectations. The assumptions made on pedestrians' expectations are crucial for the effectiveness of the method; at the same time, it highlights the adaptability of the technique to different environments and types of crowds.



(a)



(b)



(c)

Figure 6.4: (a) Time series of DTM for trajectories of one origin-destination POI pair with actual walking speed. (b) Time series of DTM with walking speed normalized and expected DTM denoted in yellow color. (c) Heat map of emotional state with valence from positive (green) to negative (red) and expected emotion (yellow).

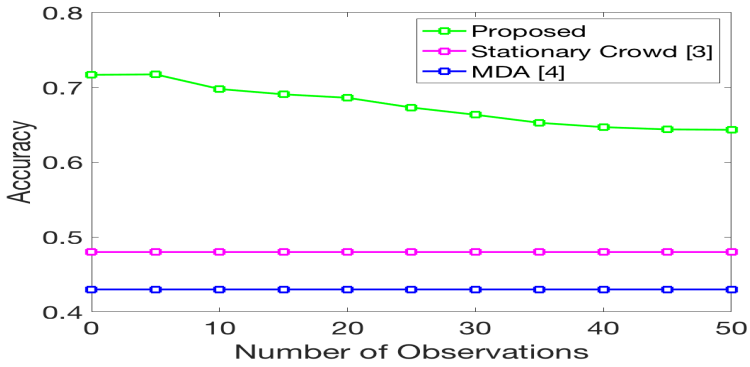


Figure 6.5: Pedestrian motivation estimation accuracy of the proposed method and comparison methods [137] and [144]. The proposed method is evaluated using diferent number of past observations, whereas comparison methods use all available observations.

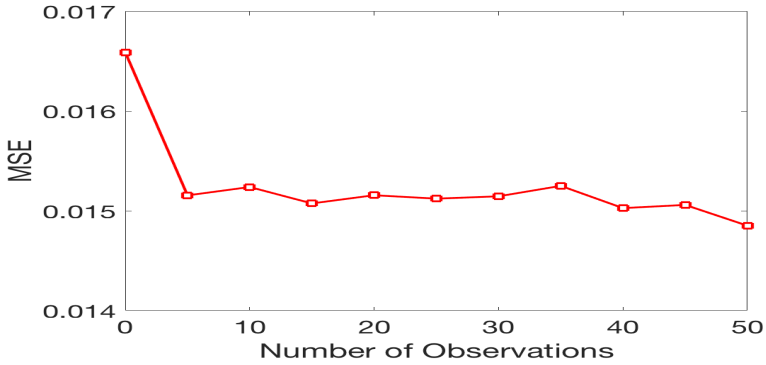


Figure 6.6: MSE of pedestrian emotion estimation evaluated using a different number of observations for pedestrian motivation estimation.

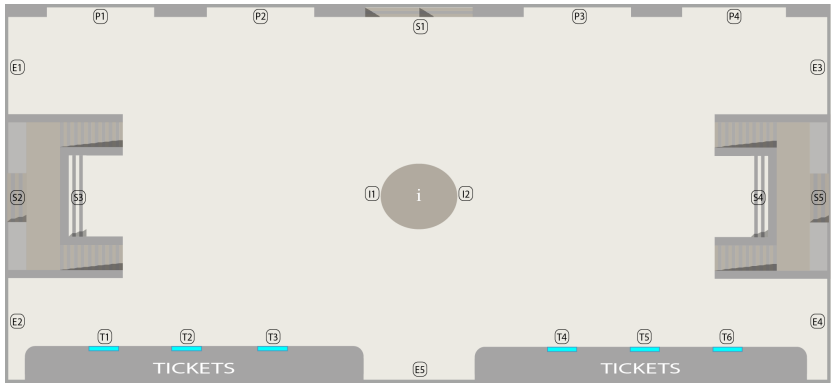
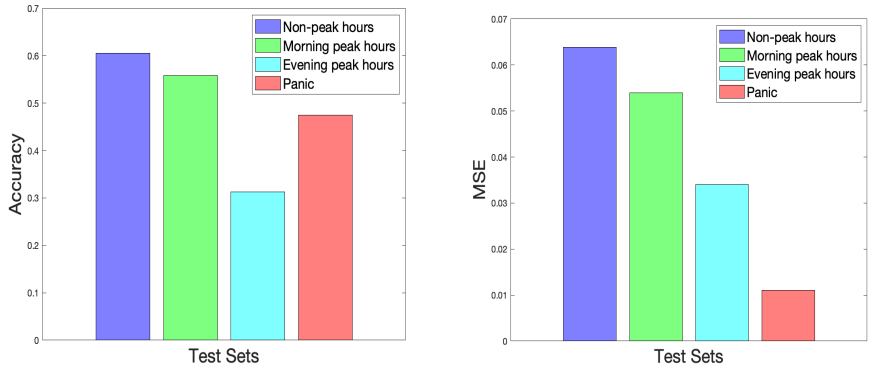


Figure 6.7: The observed environment in SC-Station dataset with POI's denoted by numerated labels.



(a) (b)

Figure 6.8: (a) Mean pedestrian motivation estimation accuracy in the four contexts presented in SC-Station dataset. (b) MSE of pedestrian emotion estimation in the four contexts presented in SC-Station dataset.

Context	Type of Crowd	Duration (seconds)	Mean/Max Pedestrian Number	Mean/Max Walking Speed (m/s)	Total Pedestrians
Non-peak hours	Ambulatory, queue	300	137 / 267	0.38 / 0.52	896
Morning peak hours	Ambulatory, queue	300	200 / 544	1.46 / 2.11	1824
Evening peak hours	Ambulatory, queue	191	706 / 3275	1.52 / 2.01	4559
Panic	Ambulatory, panic	120	129 / 325	1.19 / 2.27	498

Table 6.2: Details of SC-Station dataset.

Part III

CROWD MODEL

Chapter 7

Cyclic Behaviors in Crowds

7.1 Introduction

An essential goal of this research is to produce a crowd model capable of describe the dynamics and infer the emotion of a crowd as a whole (macroscopic level) in an analogous way to how we describe the dynamics and emotions of individual pedestrians (microscopic level) in order to retain a semantic consistency on both models. The foundation of the crowd model, as in the case of the pedestrian model, lies in the concepts of behaviors, motivations, expectations, and emotions. In this chapter, we begin to explore an appropriate way to model the dynamics of a crowd by understanding the life cycle of a crowd. To this end, we use a broad definition of the crowd as merely a group of people in proximity, independent of the existence or absence of mental unity [134]. Figure 7.1 illustrates the HDBN proposed to model the crowd. At the lower two levels and using the environment representation (SOM_p) introduced in chapter 5, we start by defining an observation and state vectors that capture the density levels in each zone of the environment at discrete time intervals. In the third level of the HDBN, the state vector is clustered into a superstate vector using a second self-organizing map (SOM_c) with the purpose to obtain a smaller set of possible states. The behavior of the crowd is conceptualized as a sequence of discrete superstate transitions over time with different sequences corresponding to a particular cycle (motivation) represented at the top level of the HDBN by a discrete latent variable. It is important to note that in this iteration of the crowd model, we do not account for expectations. Also, the emotional state

of the crowd is not inferred from its macroscopic behavior, but instead, it is a percentage representation of the discrete emotions inferred by the individual behaviors as described by the pedestrian model in chapter 5.

The remaining of this chapter is organized as follow: A comprehensive description of our proposed model is presented in section 7.2. Experiments and results to validate our model are given in section 7.3. Finally, in section 7.4, we state our conclusions and future work.

7.2 Method

This section presents the first iteration of the crowd model, as described by the HDBN in figure 7.1. In this work, we define behavior as to how an entity (pedestrian or crowd) transits among different states to achieve its motivation. For a pedestrian, a state corresponds to its location in a physical region of the environment, as we saw in previous chapters, whereas for the crowd, a state corre-

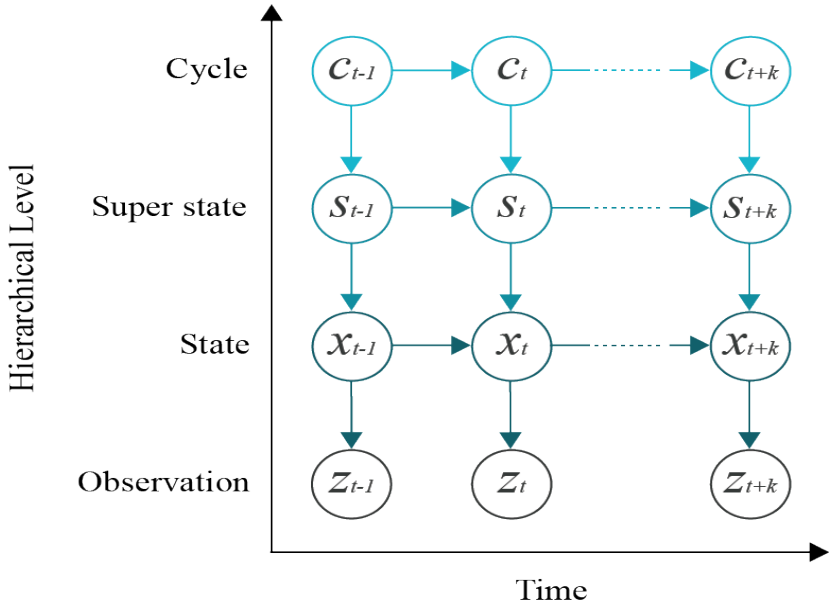


Figure 7.1: Hierarchical Bayesian networks for crowd entity.

sponds to a given configuration of people's density distribution in the observed environment. The first step is to learn the topology of the perceived environment from the trajectory of pedestrians using a self-organizing map (SOM_p) as we presented in chapter 5. An observation and state vectors capturing the density distribution in the environment are computed for the crowd, as seen in the two lower levels of the HDBN. A second self-organizing map (SOM_c) is learned to cluster state vectors into superstates, and sequences of superstate transitions are grouped into the same behavioral cycle which is described at the top of the HDBN.

7.2.1 Crowd Behavior

Supported by the work presented in [108] we argue that a crowd behaves as a collective minded entity and therefore, we can model behaviors for the crowd in a similar way to that of the pedestrian. We start our description of the crowd by defining a state vector X_t at time t as

$$X_t = \{x_{1_t}, x_{2_t}, ..., x_{n_t}; x_{i_t} \in \mathbb{R}^2\} \tag{7.1}$$

and observation vector Z_t

$$Z_t = \{z_{1_t}, z_{2_t}, ..., z_{n_t}; z_{i_t} \in \mathbb{R}^2\} \tag{7.2}$$

where x_{i_t} and z_{i_t} are the state and observation vectors of pedestrian i as defined in chapter 5, for a total of n pedestrians. Similarly, we could define $\hat{X}_t = \{\hat{x}_{i_t}\}_{i=1}^{n}$ as an estimation of the state vector of entity crowd, however the difficulty of using that definition is that $\hat{X}_t$ is prompt to irregular dimensionality between samples as pedestrians join or leave the crowd. Instead, we redefine X_t as

$$X_t = \{d_{1_t}, d_{2_t}, ..., d_{n_t}\} \tag{7.3}$$

where d_{k_t} is the estimated number of pedestrians in zone k at time t, for a total of n zones as produced by SOM_p in chapter 5. In this sense, the crowd's state vector is implicitly dependent on the detection of the pedestrian trajectories. This definition of X_t is more advantageous as it provides a vector with uniform dimensionality while maintaining meaningful information. Also, since the focus of X_t is density estimation rather than trajectory tracking, we could employ crowd density algorithms [31] [105] to achieve this task. We then proceed to model the observation and state vectors as

$$X_t = F_t x_{t-1} + B_t u_t + g_t \tag{7.4}$$

and

$$Z_t = H_t x_t + h_t \tag{7.5}$$

Where F_t is the state transition model applied to the previous state, B_t is the control-input model applied to the control vector u_t, H_t is the observation model, and g_t and h_t represent the process and observation noise, both assumed to be independent, Gaussian white. Applying an EKF over the observation and state vectors we obtain an estimation $\hat{X}_t$ and we collect the state vectors $\hat{X}_t$ into a training set

$$X_{train} = \{\hat{X}_t, \hat{X}_{t+1}, ..., \hat{X}_{t+k}; k \geq 1\} \tag{7.6}$$

and use this set to train a second self-organizing map SOM_c, that further reduces dimensionality from $\mathbb{R}^n$ to $\mathbb{R}^2$. The elements of SOM_c are as follow:

- t is the index of target input state vector $\hat{X}_t$ in the training set X_{train}.
- $\hat{X}_t$ is the target input state vector in the training set X_{train}.
- $S = \{s_1, s_2, ..., s_n\}$ is the set of neurons in SOM_c.
- $W = \{w_1, w_2, ..., w_n\}$ is the set of parametric vectors where w_k maps to neuron s_k.
- k is the index of BMU in SOM_c.

The set of nodes $S = \{s_1, s_2, ..., s_n\}$ is arranged in a hexagonal topology. Having an input data space $\mathbb{R}^n$, a parametric vector $w_k \in \mathbb{R}^n$ is learned to group all similar input vectors $\hat{X}_t$ and map them to a node s_k by finding the BMU, designated to be the node with the minimal Euclidean distance

$$\begin{gathered} ||\hat{X} - w_k|| = \min_i \{||\hat{X} - w_i||\} \\ s_k = \underset{i}{\operatorname{argmin}}\{||\hat{X} - w_i||\} \end{gathered} \tag{7.7}$$

rewritten in the form of

$$s_i = SOM_c(X) \tag{7.8}$$

where s_i is the superstate described in the third level of the HDBN. It is important to mention that SOM_c, unlike SOM_p, does not provide topological or any other semantic information by the arrangement of the nodes in S when projected to $\mathbb{R}^2$. For the crowd model we do not define discrete sequences of superstates transitions as we did for the pedestrian model in its first iteration because unlike individual pedestrians where there is a finite trajectory, under our definition of states and superstates the sequence of transitions in a crowd emerges as a cyclic process with people continuously joining and leaving the crowd. As explained at the end of this chapter, we aim to explore the cyclic behaviors of a crowd more comprehensively way in future work, but for the work presented here we describe a crowd behavior simply as a high order Markov process with a superstate transition matrix

$$P_k = \begin{pmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,n} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n,1} & p_{n,2} & \cdots & p_{n,n} \end{pmatrix} \tag{7.9}$$

where $p_{i,j}$ indicates the probability of transitioning from superstate s_i to s_j. Expressed differently, using Bayes rule we can compute the superstate transition probability as

$$\begin{aligned} p_{i,j} &= P_k(s_i|s_j) \\ &= \frac{P_k(s_j|s_i)P_k(s_i)}{P_k(s_j)} \end{aligned} \tag{7.10}$$

where P_k is a conditional probability distribution (CPD) learned from data, and k indicates the motivation (cycle) to which this CPD refers. Equation 7.10 is extended to a high order Markov process to account for more evidence to support our prediction of the next state

$$\begin{aligned} P_k(s_1, s_2, ..., s_t) &= P_k(s_t|s_1, s_2, ..., s_{t-1}) \\ &= P_k(s_t|s_1, s_2, ..., s_{t-1})P_k(s_1, s_2, ..., s_{t-1}) \\ &= P_k(s_t|s_1, s_2, ..., s_{t-1})P_k(s_{t-1}|s_1, s_2, ..., s_{t-2})P_k(s_1, s_2, ..., s_{t-2}) \\ &= ... \\ &= \prod_t P_k(s_t|s_{t-1}) \end{aligned} \tag{7.11}$$

Up to this iteration of the crowd model, we focus on the transition of superstates and do not account for the temporal component of these transitions.

7.2.2 Crowd Motivation, Expectation and Emotion

The definition of behavior in the previous section points to the sequence of superstate transitions to most likely be observed in the crowd when a particular cyclical behavior is observed. We extend this definition to account for multiple cycles where each cycle corresponds to a different motivation. Having our training data X_{train} manually separated into subsets for each cycle, we can proceed to learn the same models of equations 7.10 and 7.11 for n cycles. Furthermore, we can make use of a Bayes classifier to infer the motivation C_t of the crowd based on previously observed superstate transitions

$$C_t = \underset{k}{\operatorname{argmin}} \prod_{t=1}^{n} Pr(k)P_k(s_t|s_{t-1}) \tag{7.12}$$

where $Pr(k)$ represents the a priori probability of motivation k, n is the total number of learned cyclic behaviors, and $C_t = k$ indicates the motivation of the crowd at time t. Given that multiple motivations may be learned from data, we define the transition of motivations C_t to C_{t+1} by a discrete-time Markov chain

$$\begin{aligned} P(C_{t+1}|C_1, C_2, ..., C_t) &= P(C_{t+1}|C_t) \\ &= \frac{P(C_t|C_{t+1})P(C_{t+1})}{P(C_t)} \end{aligned} \tag{7.13}$$

holding the Markov property that the next motivation C_{t+1} depends only on the current motivation C_t. The description of the crowd model concludes here, leaving the concepts of expectation and emotions to be addressed in the following chapters.

7.3 Experiments and Results

Two experiments are conducted to evaluate the proposed crowd model: the first experiment uses a dataset generated from a simulation tool [28] [29] and focuses on evaluating the model's capability to predict future states of the crowd within the same cyclic behavior, and the second experiment employs a dataset containing real-world observations of pedestrians [144] and aims to assess the inference of crowd motivations by identifying the cyclic behavior model that best describes the dynamics in the crowd.

7.3.1 Experiment 1: Synthetic Dataset

The experiment described in this section employs the same dataset used in chapter 5, and its details are summarized in table 7.1. The dataset containing pedestrian trajectories is divided into a training set T_{train} and a testing set T_{test} with equal proportions.

	Training dataset	Test dataset
Duration (hours)	5	5
Positive trajectories	978	894
Normal trajectories	1770	1815
Negative trajectories	252	291
Total trajectories	3000	3000

Table 7.1: Details of training and testing datasets produced from simulations.

Model Training

We start by learning two SOM's: SOM_p for the topological representation and partition of the observed environment and SOM_c to cluster crowd states into superstates and model the behavior of the crowd as a whole. Both self-organizing maps SOM_p and SOM_c are initialized with similar parameters. The set of neurons on each SOM contains 100 neurons (10 rows and 10 columns) is initialized with random weights and in a hexagonal arrangement spread across the corresponding input space. Distance between neurons is calculated by the number of links among them. The initial neighborhood size is 3, with 100 steps for the ordering phase. The training phase is done over 500 epochs by competitive layer

but without bias, updating the winning neuron and all other neurons within the given neighborhood using Kohonen rule.

The topological representation of the environment learned by SOM_p using the training dataset T_{train} is presented in figure **??**. Using the description provided by SOM_c, the sets Z_{train} and X_{train} for the observation and state vectors are computed. Next, the collection of estimated state vectors X_{train} is used to train SOM_c, enabling us to obtain the set S_{train} containing superstate vectors, and learn the probabilistic models to describe the crowd's cyclic behavior and motivation.

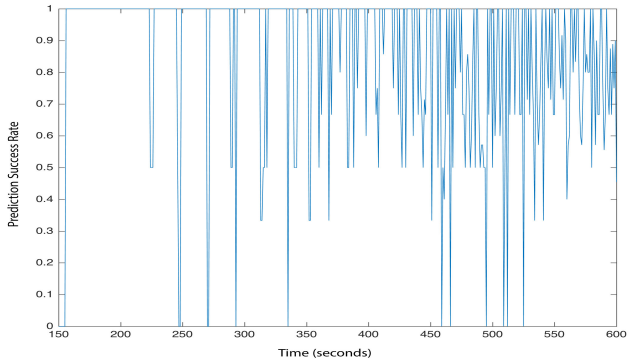
It is crucial to notice that unlike SOM_p, a plot of SOM_c does not provide a visual semantic due to the high dimensionality of the state vectors, so an illustration of the resulting SOM_c is not provided.

Model Evaluation

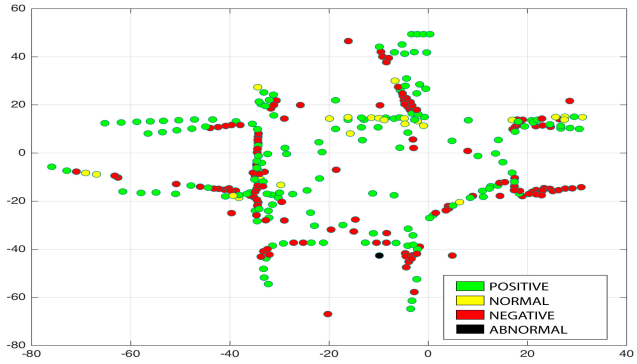
Making use of T_{test} and SOM_p, we obtain the testing set X_{test} to be employed in evaluating the model's ability to predict the behavior of the crowd, that is, the next state and transitions among states of SOM_c.

In figure 7.3a, we can observe the behavior prediction accuracy for the crowd oscillating more consistently between 50% and 94%, with a mean accuracy of 81%. As the data used in this experiment correspond to only one type of cyclic behavior, no estimation of the crowd's motivation is produced but will be addressed in future chapters. Also, the association between crowd's behavior/motivation and emotion is not defined yet, so we limit to represent the emotional state of the crowd as a summary of the individual emotions inferred by the pedestrian model.

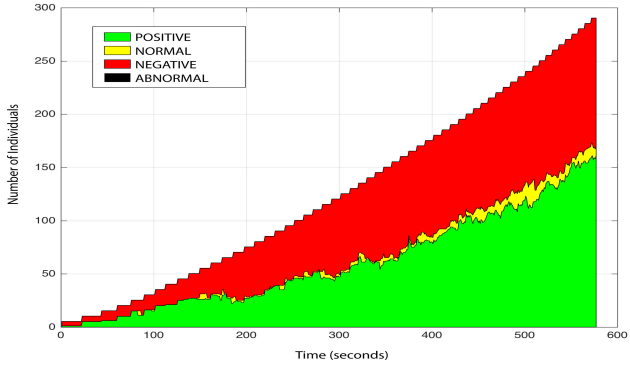
In sub-figure 7.3b we present a picture of pedestrian emotion estimations produced by the pedestrian model introduced in chapter 5. Sub-figure 7.3c displays the summary of inferred pedestrian emotions in an observation period of 600 seconds, during which a positive emotion becomes predominant as the number of pedestrians joining the crowd increases.



(a)



(b)



(c)

Figure 7.3: (a) Crowd behavior prediction accuracy. (b) Picture of online pedestrian emotion estimation. (c) Picture of online crowd emotion as a summary of pedestrian emotions.

7.3.2 Experiment 2: C-Station Dataset

In the second experiment, we employ the Grand Central Station dataset [144], which provides manual annotations of 12,684 observed pedestrian trajectories from a one-hour crowd surveillance video. The dataset is divided into a training set T_{train} containing 70% of trajectories and a testing set T_{test} containing the remaining 30% of trajectories.

Model Training

The environment is learned using T_{train} to train the topological map SOM_p, configured in a hexagonal arrangement with 50 neurons (10 columns and 5 rows) and neurons weights are initialized randomly within the input space, and the initial neighborhood size is 3 with 100 steps in the ordering phase. The training phase is performed over 500 epochs by competitive learning without bias. The resulting topological map SOM_p is shown in figure 7.4.

The sets T_{train}, and T_{test} containing pedestrian trajectories are used in combination with SOM_p to compute the collection of state vectors X_{train} and X_{test}. The set X_{train} is then used to train SOM_c. This SOM is configured in a hexagonal arrangement with 100 neurons (10 columns and 10 rows), the weights of the neurons are initialized randomly across the input space, and the initial neighborhood size is 3 with 100 steps in for ordering phase.

The training phase is performed over 500 epochs by competitive layer without bias. Making use of SOM_c, we can proceed to convert the sets of state vectors X_{train} and X_{test} into sets of superstates S_{train} and S_{test}.

A total of seven different crowd behaviors are manually labeled and separated into training and testing sets, $S^1_{train}, S^2_{train}, ..., S^7_{train}$ and $S^1_{test}, S^2_{test}, ..., S^7_{test}$.

Model Evaluation

The criteria in identifying the different crowd behaviors is characterized by grouping all detected pedestrians heading to the same direction, in the experiments presented here we employ seven different directions: north-east, north, north-west, south-west, south, south-east, and mixed directions. Additionally, three variations of each testing set were used for evaluation where 100%, 75% and 50% of people act according to the testing set's intended behavior and the remaining percentage of people act in different behaviors. A plot of each behavior is presented in figure 7.5.

The performance of the models' capability to identify each behavior is recorded after various periods of observation: 1 observation, 10 observations, 20 observations, 30 observations. Results are shown in table 7.2, separated into 3 sub-tables for each variation of the testing sets, by rows for each behavior and by columns for each length of the observation period.

The results in table 7.2a show a low accuracy to identify most behaviors after only one observation but consistently increase after more observations; this is reasonable as 100% of people in the testing set moving in the same behavior is an improbable scenario from the provided data. However, in table 7.2b the results after just one observation are high as only 75% of people follow the detected behavior, which is a more typical case.

Results in table 7.2c where only 50% of people follow the intended behavior are still high for most of the behaviors. A testing set with less than 50% is not included in the experiment because such a low number of people following the same behavior would imply that this is not the dominant behavior in the crowd.

A direct comparison between our crowd model and other related methods would be inaccurate since our method models the predominant behaviors of the crowd as a whole whereas existing methods model one or several behaviors based on partial features of the crowd. However, the next chapter will address a congruent comparison with relevant models. Additionally, a significant portion of literature that studies crowd behavior is focused on abnormality detection, which is not covered in the experiments presented here.

It is also relevant to notice that the conducted experiments made use of the ground truth provided for this dataset; therefore the results displayed do not account for the accuracy error added by the underlying algorithm employed for pedestrian tracking and detection.

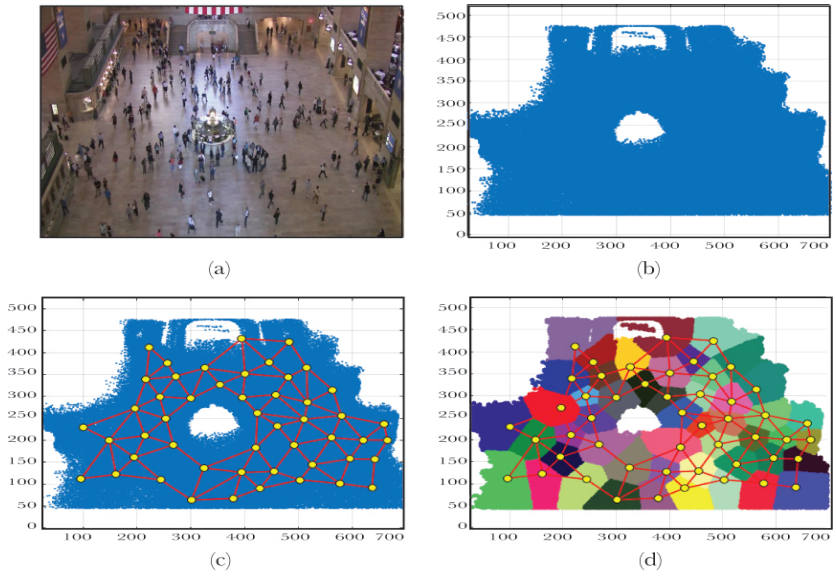


Figure 7.4: (a) Footage of New York Grand Central station. (b) Data points of annotated trajectories. (c) Topology learned with growing neural gas networks. (d) Environment representation divided by regions.

Behavior	$t-1$	$t-10$	$t-20$	$t-30$
North-east	0.0560	0.3098	0.4416	0.5337
North	0.0815	0.2732	0.3891	0.4951
North-west	0.5423	0.6259	0.6477	0.6531
South-west	0.0325	0.2006	0.2860	0.3316
South	0.5055	0.5685	0.6446	0.7331
South-east	0.2890	0.4620	0.5689	0.6539
Mixed	0.9600	0.9954	1.0000	1.0000

(a) Test set with 100% of trajectories moving according to each behavior.

Behavior	$t-1$	$t-10$	$t-20$	$t-30$
North-east	0.9730	0.9984	1.0000	1.0000
North	0.9769	0.9989	1.0000	1.0000
North-west	0.9775	0.9969	1.0000	1.0000
South-west	0.9309	0.9914	1.0000	1.0000
South	0.9955	1.0000	1.0000	1.0000
South-east	0.9845	0.9849	0.9934	0.9969
Mixed	0.8980	0.9869	1.0000	1.0000

(b) Test set with 75% of trajectories moving according to each behavior.

Behavior	$t-1$	$t-10$	$t-20$	$t-30$
North-east	0.9870	0.9984	1.0000	1.0000
North	0.9760	0.9989	1.0000	1.0000
North-west	0.9775	0.9969	1.0000	1.0000
South-west	0.9309	0.9914	0.9994	1.0000
South	0.9901	1.0000	1.0000	1.0000
South-east	0.3310	0.3761	0.4406	0.4550
Mixed	0.6620	0.6238	0.5593	0.5449

(c) Test set with 50% of trajectories moving according to each behavior.

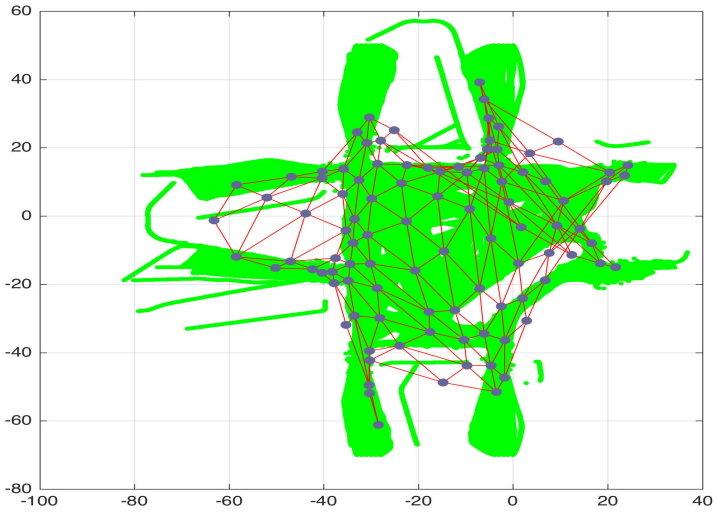
Table 7.2: Results of model's performance to identify learned behaviors after 1, 10, 20, and 30 observations. Three different test sets are employed where 100%, 75%, and 50% of the trajectories move according to each behavior.

7.4 Conclusions

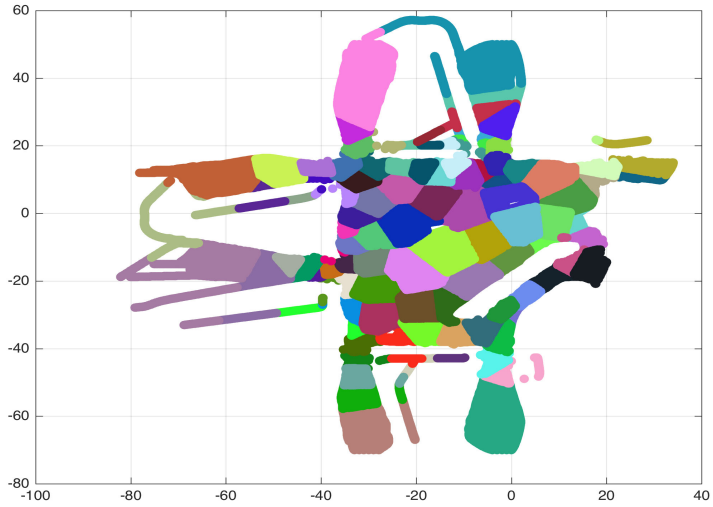
In this chapter, we presented the first iteration of the crowd model aiming to explore a suitable approach to describe the dynamics of a crowd as a whole, under realistic crowded environments. Behaviors are modeled by changes in density distribution in the crowd rather than by observing single trajectories of pedestrians. This approach is advantageous when crowd density is high as in such circumstances, people counting methods perform better than pedestrian tracking techniques. Also, seen the crowd as a whole provides a more comprehensive understanding of the crowd's dynamics. The conducted experiments support the viability of achieving a macroscopic view of the crowd yielding consistently reliable predictions of its future behavior and identification of motivation. In particular, the crowd model performs with high accuracy to identify different behaviors in crowds even when the predominant behavior is exhibited by as low as only half of the pedestrians in the crowd.

Our approach accounts for scenarios with multiple origin and destination points, and it explores the idea of the crowd as a separate entity with its cyclic behavior and motivation. The overall hypothesis is that a crowd can be described as a separate entity with its own behavior and motivations in a way that is consistent with those of the individual pedestrians, as suggested by [41]. In this particular case, we have a rather simple model that treats the crowd emotion as a sum of the emotions of the pedestrians in the crowd. The approach presented here is applicable to crowded real-life environments for monitoring automation intended to identify and prevent dangerous situations as well as to improve crowd control. Furthermore, contributions of this nature are essential for the development of robust cognitive dynamic systems intended for smart cities.

The next iteration of the crowd model will focus on extending the model to consider a proper definition of expectations and emotions for the crowd and their association to the already defined behaviors and motivations. We will also explore the interaction of pedestrian and crowd emotions, enabling us to better understand causality and contagion of emotional states among pedestrians and its impact in the crowd as a whole.



(a)



(b)

Figure 7.2: (a) Training data (green) and the self-organizing map SOM_p (red edges and blue nodes). (b) Environment partitioned into zones with colors assigned randomly.

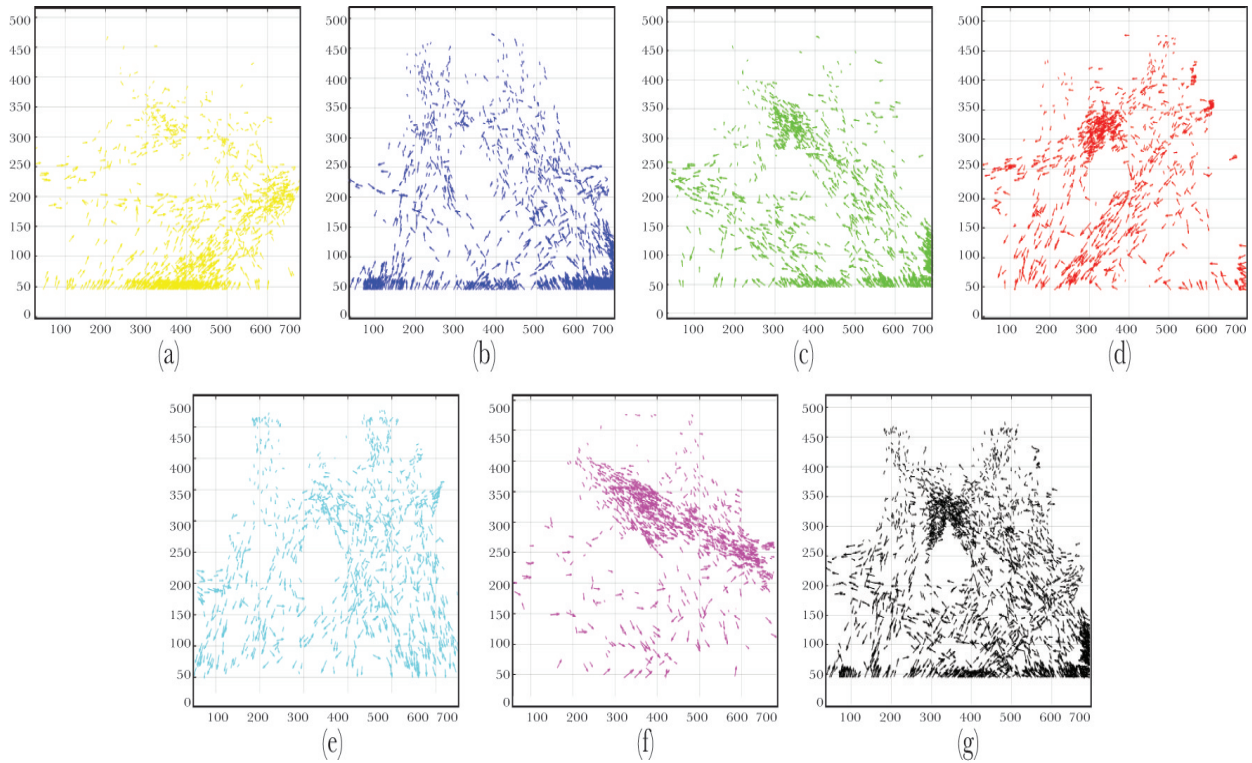


Figure 7.5: List of behaviors learned from the training dataset. (a) Trajectories heading north-east. (b) Trajectories heading north. (c) Trajectories heading north-west. (d) Trajectories heading south-west. (e) Trajectories heading south. (f) Trajectories heading south-east. (g) Trajectories heading to multiple directions.

Chapter 8

Crowds Within Crowds: A Sub-Crowd Model for Emotion Estimation

8.1 Introduction

An essential facet in the study of crowds lies in understanding the influence among pedestrians [137] [143]. One relevant study in this area examines the formation of groups in crowds [73] confirming the intuition that pedestrians in closer proximity exert a stronger influence than those in a more distant location, leading to the formation of crowds within crowds. In this chapter, we turn our attention to this fact, crowds within crowds, and work in enabling the previously presented crowd model to adequately describe sub-crowds inheriting the core hypothesis of using behaviors, motivations, and expectations in deriving emotional states for each sub-crowd in the environment. The previous iteration of the crowd model provided a starting point in showing the viability to represent the crowd at a macroscopic level and devise a way to model its behavior as a transition of states dictated by an associated motivation. Preserving the topological representation previously introduced, we switch from the idea of one crowd to multiple sub-crowds in the same location, where each sub-crowd is comprised by the pedestrians present in the same subregion of the environment. Also, this iteration improves the selection of features used to represent

the state of the sub-crowd and its learned behavior model. The crowd model presented in this chapter is based on the implementation of a HDBN as shown in figure 8.1, with one instance of the HDBN for each sub-crowd as associated with a particular subregion of the environment. Furthermore, this version of the crowd's HDBN, as it was in the case of the final version of the pedestrian HDBN, is reduced to its most essential components captured in three hierarchical levels: two continuous-valued observation and state vectors to describe the state of the sub-crowd based on the PTF measured in its corresponding subregion, and a discrete-valued variable to identify the sub-crowd's motivation as associated to the behavior model that best describes the observed sequence of states. Also, similar to the pedestrian model, the components of expectation and emotion are described outside the HDBN and inferred based on sub-crowd's behavior and motivation. The expectation of a sub-crowd is quantified as the area under the curve of the predicted PTF according to the behavior model associated with the estimated motivation. The emotion of a sub-crowd is conceptualized as a valence-based continuous value derived from the breach between expected and observed PTF. Two experiments were conducted to evaluate in great detail the viability of the proposed crowd model. The first experiment makes use of real-world data to assess the feasibility and accuracy of our model to describe sub-crowd behaviors and infer its consequent motivations, expectations, and emotions. The second experiment simulates the environment observed in the first experiment, and generates synthetic data to allow for broader variation in conditions, enabling us to measure how well the crowd model adapts to different contexts. Both experiments yielded positive results, confirming the viability of the crowd model to provide a macroscopic representation of the sub-crowd's behavior while producing an inference of the sub-crowd's emotion that is consistent with that of the pedestrians comprising it. For the remaining content of this chapter, in section 8.2 we provide an in-depth description of the proposed method. Section 8.3 presents two experiments conducted to evaluate the effectiveness of the proposed method. Finally, section 8.4 postulates the conclusions drawn.

8.2 Method

This second iteration of the crowd model aims to represent the conglomerate of pedestrians not as a single crowd as it was the case of the first iteration of the crowd model, but as a collection of sub-crowds in the environment based on spatial proximity among pedestrians. The model follows the implementation of

the HDBN shown in figure 8.1 and one instance of this HDBN is employed to describe each particular sub-crowd. The topology of the environment is learned from data and divided into subregions in an unsupervised way as we did in the previous chapter. The observation and state vectors Z_t and X_t of sub-crowd i represented by the two lower levels of the HDBN are computed every time instance t by measuring the PTF of the pedestrians present in the subregion i. The behavior model k to predict the transition among state vectors is selected according to the inferred motivation $m_t = k$. Finally, the expectation is quantified and used to infer the emotional state of the sub-crowd. In this direction, behavior models, motivations and expectations are learned and estimated for each subregion, resulting in a single emotion estimation representative of the emotion experienced by the members of that sub-crowd. The crowd model relies on the detection and partial tracking of pedestrians as extracted from surveillance cameras as the initial input signal.

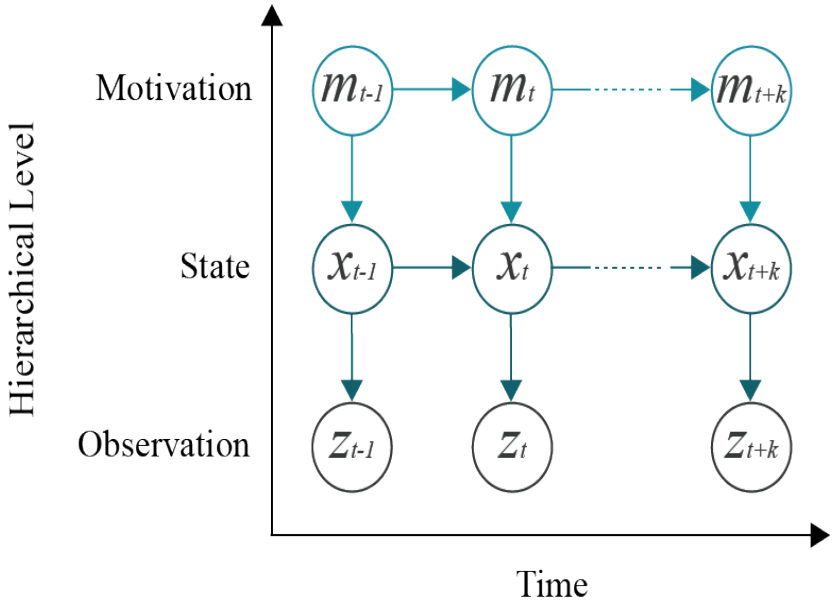


Figure 8.1: HDBN for a sub-crowd entity.

8.2.1 Sub-Crowd Representation

The proposed method aims to split the environment into subregions and compute representative features of the dynamics of the crowd for each subregion. To divide the environment into subregions, the SOM [72] is a suitable method as it learns in an unsupervised way to cluster samples from the input space into a particular neuron, but unlike other clustering methods, it uses a neighborhood function to maintain the topological properties of the input space. The SOM is defined by a set of neurons with a weight vector $W = \{w_1, ..., w_n\}$ where $w_i \in \mathbb{R}^2$ is the position of neuron i for a total of n neurons, and a function $Q(x) = i$ mapping each sample $x \in \mathbb{R}^2$ from the input space into a neuron i. The training phase of a SOM uses competitive learning to adjust the weights W of its neurons based on a training set $x_{train} = \{x_1, ..., x_k\}$ where $x_i \in \mathbb{R}^2$ is the detected position of a pedestrian i. An illustration of this process is provided in figure 8.3. Given the scope of this method is on ambulatory crowds, the PTF feature proposed by [56] is a suitable measure to describe the crowd dynamics as it relates crowd density with movement. Furthermore, the research done by Fruin [56] introduces the LOS standards to asses the quality of PTF based on the nature of the environment (e.g., walkways, stairways, queuing areas) and the flow orientation (e.g., unidirectional bidirectional, multi-directional). The LOS values have been adjusted for the specific characteristics of the environment studied in this work and are presented in table 8.1. The state of the sub-crowd i at time t is described by a vector state

$$X_t = \frac{s_t}{a_t} \tag{8.1}$$

where X_t is the PTF measured in pedestrians per meter per minute $(pr/m/min)$, s_t is the average pedestrian walking speed (m/min), and a_t is the average area per pedestrian within the traffic stream (m^2/pr). The value of a is calculated by the time-space (TS) method [56] as follow

$$a = \frac{TSsupply}{TSdemand} = \frac{TS}{nt} \tag{8.2}$$

where T is the time of the analysis period (min), S is the effective net area of the analysis space (m^2), n is the number of pedestrians currently occupying the space, and t is the predicted occupancy time of pedestrians. The observation vector is given by

$$Z_t = H_t X_t + v_t \tag{8.3}$$

LOS	PTF pr/m/min	Description
A	0 - 23	Threshold of free flow, convenient passing, conflicts avoidable.
B	23 - 33	Minor conflicts, passing, and speed restrictions.
C	33 - 49	Crowded but fluid movement, passing restricted, cross, and reverse flows difficult.
D	49 - 66	Significant conflicts, passing and speed restrictions, intermittent shuffling.
E	66 - 82	Shuffling walk reverse, passing and cross flows very difficult, intermittent stopping.
F	82 - Max	Critical density, flow sporadic, frequent stops, contact with others.

Table 8.1: Description of LOS. The PTF is measured in pedestrians per meter per minute $(pr/m/min)$.

where H_t is the observation model mapping the state space into the observed space, and v_t is the observation noise assumed to be independent, Gaussian white. Computing the PTF requires the detection and partial tracking of a pedestrian for which state of the art methods perform well [141]. Details on the detection and tracking of the pedestrians is out of the scope of this work and will not be further discussed.

8.2.2 Sub-Crowd Behavior

A sequence of observations of the PTF of a subregion is captured as a discrete set of data points ordered in time, which fluctuates under different thresholds according to the context of the situation. Because of the nature of this data, modeling the dynamics of PTF in a subregion is addressed as a time series regression problem, and a Gaussian process regression model is used:

$$X_{t+1} = h(X_{1:t})^T \beta + f(X_{1:t}) + u_k \tag{8.4}$$

where X_{t+1} is the PTF of subregion i at time $t+1$, the input vector $X_{1:t} =$

$\{X_1, X_2, ..., X_{t-1}, X_t\}^T$ is a sequence of past measurements with an observation window size of t data points, $h(X_{1:t})$ is a set of basis functions, β is a vector of basis function coefficients, $f(X_{1:t}) \sim GP(0, k(X_{1:t}, X_{1:t}^T))$ is a set of latent variables $f(X_t)$ from a Gaussian process with zero mean and covariance function $k(X_{1:t}, X_{1:t}^T)$, and u_k represents the noise assumed to be Gaussian white noise with covariance $Q_t : u_t \sim \mathcal{N}(0, Q_t)$. A one-step-ahead prediction $\hat{X}_{t+1}$ is computed by

$$\hat{X}_{t+1} = P(X_{t+1}|f(X_{1:t}), X_{1:t}) \tag{8.5}$$

where

$$P(X_{t+1}|f(X_{1:t}), X_{1:t}) \sim \mathcal{N}(X_{t+1}|h(X_{1:t})^T\beta + f(X_{1:t}), \sigma^2) \tag{8.6}$$

And further n-step-ahead predictions

$$\hat{X}_{t+1:t+n} = \{\hat{X}_{t+1}, \hat{X}_{t+2}, ..., \hat{X}_{t+n}\} \tag{8.7}$$

are computed recursively adding the previous predictions to the input vector. To asses the accuracy of predictions, a regression error function is defined

$$L(X_{t:t+n}, \hat{X}_{t:t+n}) = \frac{1}{n}\sum_{k=t}^{n}(X_k - \hat{X}_k)^2 \tag{8.8}$$

where $L(X_{t:t+n}, \hat{X}_{t:t+n})$ computes the MSE between a set of observations $X_{t:t+n}$ and predictions $\hat{X}_{t:t+n}$, both corresponding to the same interval of time $t, t+1, ..., t+n$.

8.2.3 Sub-Crowd Motivation

Following the idea of analyzing a crowd by subregions, the pedestrians present in a subregion have the motivation to reach a (partial or final) destination within that same subregion, for which they follow a walking behavior. The combination of these observed walking behaviors produces a particular PTF pattern under different contexts, for which the motivation of the crowd at that subregion is to maintain a PTF consistent to the demand of the pedestrians in it. This PTF pattern may change depending on the context of the situation; hence, this work proposes that a subregion has a different motivation for each context. To illustrate the concept of contexts, consider the flow of passengers in a train station, which is different comparing morning rush hours to quiet Sunday afternoons, and even more to unusual situations like panic scenarios.

A sub-crowd behavior model is learned for each manually identified context as described in the previous section and associated to a particular motivation. For a set of sub-crowd behavior models $1, 2, ..., n$, the motivation of subregion i at time t is

$$m_t^i = k \tag{8.9}$$

if the model k best describes the observed PTF. As multiple models exist for the same subregion i, estimating the motivation m_t^i is treated as a multi-class classification problem and a one-versus-all approach is used

$$\hat{m}_t^i = \underset{k \in \{1,...,n\}}{\operatorname{argmin}} L(X^i, \hat{X}_k^i) \tag{8.10}$$

where X^i is a set of PTF observations in subregion i, $\hat{X}_k^i$ is the PTF prediction produced by the model k as defined in equation 8.6, for all n models, and L computes the regression error function between observed and predicted values as stated in equation 8.8.

8.2.4 Sub-Crowd Expectation

In the presence of uncertainty, expectation comes from our assessment of the most likely outcome. In the context of a sub-crowd, once a motivation $\hat{m}_t^i = k$ is estimated, expectation is conceptualized as the expected PTF corresponding to the predictions $\hat{X}_{t:t+n}^i$ produced by model k. The expectation of sub-crowd i is quantified using the trapezoidal rule over the curve depicted by the data points $\hat{X}_{t:t+n}$

$$AUC_{exp}^i = \sum_{t=1}^{n-1} \frac{\hat{X}_t + \hat{X}_{t+1}}{2} \Delta_t \tag{8.11}$$

where

$$|\hat{X}_{t:t+n}^i| = n \tag{8.12}$$

and Δ_t is the time interval between data points. The actually observed PTF captured in $X_{t:t+n}$ is quantified in the same way

$$AUC_{act}^i = \sum_{t=1}^{n-1} \frac{X_t + X_{t+1}}{2} \Delta_t \tag{8.13}$$

The quantification of expected and actual PTF for a specific sub-crowd provides a simplified way to measure the deviation from expected behavior and will be used in the next step of the crowd model.

8.2.5 Sub-Crowd Emotion

The proposed description of emotion for sub-crowd i at time t is a continuous value E_t^i, intended to be a reflection of the emotions experienced by the pedestrians occupying subregion i at time t. Consistent with how pedestrian emotions are presented in chapter 6, the value E_t^i indicates a point in the valence axis restricted to the range from 0 (negative) to 1 (positive). Following a similar approach to the one used to estimate the emotion of pedestrians, this method hypothesizes the emotion of sub-crowd i is determined by the deviation between the expected and observed PTF. Using the values AUC_{exp}^i and AUC_{act}^i, the deviation from expected behavior is measured by

$$AUC_{\Delta}^i = \frac{AUC_{exp}^i - AUC_{act}^i}{AUC_{exp}^i} \tag{8.14}$$

and the emotion of sub-crowd i at time t is computed as

$$\begin{aligned} E_t^i &= E_{exp}^i + E_{exp}^i f(AUC_{\Delta}^i) \\ &= E_{exp}^i (1 + f(AUC_{\Delta}^i)) \end{aligned} \tag{8.15}$$

Where E_{exp}^i is the expected emotion manually associated with the behavior model k of sub-crowd i if $m_t^i = k$, and $f(x)$ is a standard logistic regression function with parameters $k = 1$, $x_0 = 0$, and $L = 1$

$$\begin{aligned} f(x) &= \frac{L}{1 + e^{-k(x - x_0)}} \\ &= \frac{1}{1 + e^{-x}} \\ &= \frac{e^x}{e^x + 1} \end{aligned} \tag{8.16}$$

Analogous to the estimation of pedestrian emotions, this way of computing crowd emotions relies on a correct estimation of the motivation m_t^i, and it provides the same advantage that prevents previous wrong motivation predictions from influencing future estimations of sub-crowd emotions. Given that E_t^i represents the emotion of sub-crowd i at time t, the collection of all sub-crowd emotions can be collected into a single vector

$$E_t = \{E_t^1, E_t^2, ..., E_t^n\} \tag{8.17}$$

where n is the total number of sub-crowds, one for each subregion in the environment. Representing crowd emotions not as a single value but as a vector provides a more granular understanding of the state of the crowd and potential crowds within it without losing a macroscopic view of the crowd as a whole. This consideration allows our model to adapt well in either the presence or absence of psychological unity [66] and collective emotions [68].

8.3 Experiments

The effectiveness of the proposed crowd model is demonstrated by means of two experiments: the first experiment makes use of real-world data to evaluate the ability of our method to learn sub-crowd behavior models and produce emotion estimations in a single context, and the second experiment employs a more extensive dataset produced from simulations to evaluate our method's ability to identify sub-crowd motivations among multiple observed contexts and infer its emotional state.

8.3.1 Experiment 1: C-Station Dataset

The first experiment is conducted using the C-Station dataset [137], which contains annotation of pedestrian trajectories from footage of New York Grand Central Station. Details of this dataset are summarized in table 6.1 of chapter 6. Given the small variation in pedestrian density and walking speed, the entirety of the dataset is identified as a single context, and only one motivation is labeled for each subregion of the environment. The goal of this experiment is to test the capability of our method to learn the sub-crowd behaviors and use them in predicting expected PTF to infer the emotional state. To asses how representative the sub-crowd emotion is from the pedestrians conforming it, we compare the inferred emotional state of the sub-crowd against the mean of pedestrian

emotion labels produced in chapter 6. A 70% of the trajectories is assigned to the training set D_{train} and 30% to the testing set D_{test}

Model Training

For the training phase, the topology of the environment is learned utilizing a SOM using the trajectories in the training set D_{train}. The decision on the number of subregions to use depends on the desired level of granularity to describe the crowd at a macroscopic level. For this dataset, a SOM with 25 subregions is selected, and the resulting topology can be observed in figure 8.3. The training set D_{train} is used to compute a set of state vectors X_{train} as prescribed in sub-section 8.2.1 and a crowd behavior model is learned for each subregion i using its PTF sequence X^i_{train} as defined in equation 8.4. The training phase is concluded by manually assigning the expected emotion E^i_{exp} to the crowd behavior model of each subregion i.

Model Evaluation

Evaluation on the testing set D_{test} starts by computing the PTF vectors X_{test} and then calculating the estimation of crowd motivations $\hat{m}^i_t$ for each subregion i. However, given that only one kind of scenario (i.e., context) is observed throughout the dataset, only one crowd behavior model is learned for each subregion, and therefore it becomes irrelevant to evaluate the accuracy in predicting m^i_t to select the right crowd behavior model.

Employing the values of AUC^i_{exp} and AUC^i_{act} calculated as explained in subsection 8.2.4, the emotion estimation $\hat{E}^i_t$ for each subregion i is computed over the observation period in the testing set. The accuracy of the crowd model to estimate crowd emotions is measured by the MSE between $\hat{E}^i_t$ and E^i_t, where E^i_t represents the mean emotion ground-truth of the pedestrians present in subregion i at time t as it was computed in chapter 6. The MSE for each subregion i is presented in figure 8.2, showing a mean MSE of 0.0218 and standard deviation of 0.004. The resulting MSE values are consistently low in all subregions, supporting the idea that the proposed conceptualization of crowd emotion indeed reflects the emotion of the pedestrians.

8.3.2 Experiment 2: Grand Central Station Synthetic Dataset

This experiment aims to assess the performance of our model under a broader range of conditions. For this reason, we virtually replicate the observed envi-

ronment in the first experiment and a synthetic dataset is generated using the agent-based social force model presented in chapter 4 with details of the produced data are presented in table 6.2 of chapter 6. The dataset is divided accordingly to 4 manually labeled contexts: non-peak hours, morning peak hours, evening peak hours, and panic. For each context k, the data points are separated into a training set D_{train_k} (70%)and testing set D_{test_k} (30%).

Model Training

The training phase starts with learning the environment's topology by feeding the annotated trajectories of all training sets $D_{train_1}, ..., D_{train_4}$ to a SOM configured to 25 neurons (subregions), with the result of this process illustrated in figure 8.4. For each training set, the PTF state vectors $X^i_{train_k}$ in each subregion i are computed and used to learn a Gaussian process model k capable of predicting PTF behavior within that same subregion i. Finally, the crowd emotion labels for each subregion are generated by computing the mean pedestrian emotion label generated in the second experiment of chapter 6.

Model Evaluation

For the testing phase, identifying the current context for subregion i is done by predicting the sub-crowd motivation m^i_t using the previously observed PTF

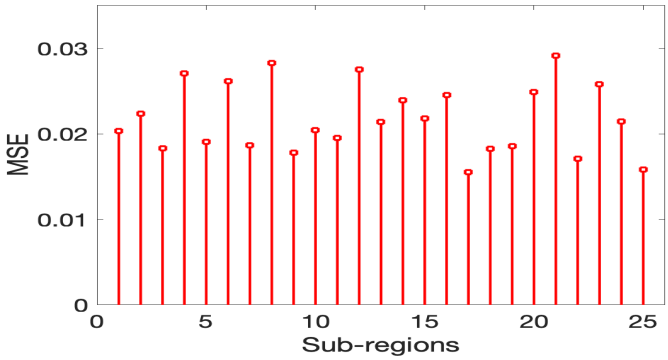


Figure 8.2: MSE between crowd emotion estimation and mean pedestrian emotion for each subregion computed using C-Station test set.

sequence $X^i_{t-N:t}$ with a minimum observation period of $N = 30$. The motivation prediction accuracy for every subregion in each test set is presented in figure 8.5.

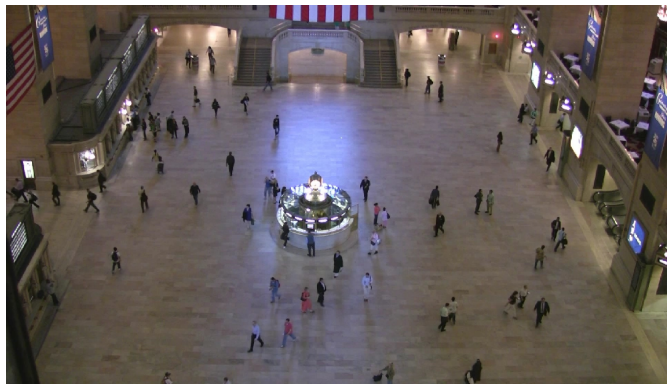
Crowd motivation prediction results for each context are broadly similar to those for pedestrian motivation prediction, with non-peak hours producing the highest accuracy of 0.95, morning and evening peak hours with an accuracy of 0.47 and 0.55, respectively and finally panic context with 0.42. A correlation between pedestrian and crowd motivations is not required as one is not an aggregation of the other. However, comparing the pedestrian motivation prediction results of the second experiment of chapter 6 with the sub-crowd motivation prediction we find a consistency that is informative in confirming the level of predictability by which pedestrians move on different contexts at a microscopic and macroscopic level. The predicted sub-crowd motivation $\hat{m}^i_t = k$ allows the selection of the appropriate Gaussian model k to be used in forecasting the expected subsequent PTF sequence $\hat{X}^i_{t:t+N}$. The crowd emotion E^i_{t+N} for each subregion i is inferred from the deviation between expected $\hat{X}^i_{t:t+N}$ and actually observed $X^i_{t:t+N}$ sequences as captured by AUC^i_{exp} and AUC^i_{act}.

Figure 8.6 shows the MSE between the inferred crowd emotion and the mean pedestrian emotion label of people present in that same subregion evaluated for all contexts. The panic scenario consistently yields the smallest MSE values in all subregions, whereas the MSE in the remaining scenarios differs significantly more between subregions. Low MSE values indicate a strong correlation between pedestrian and crowd emotions, confirming that the proposed representation of crowd emotion is indeed representative of the pedestrian emotions, although to different degrees in every scenario.

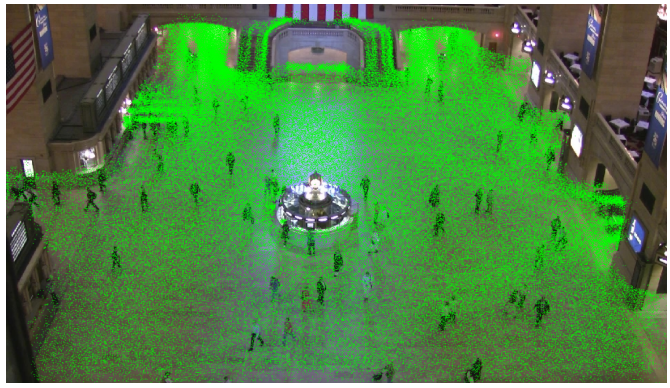
8.4 Conclusions

This chapter presented a second iteration of the crowd model first introduced in chapter 7. The model was extended to account for crowds within crowds and provided a formal definition of behavior, motivation, expectation, and emotion at the sub-crowd level that is analogous to the pedestrian model. The intention to model both the pedestrian and the crowd analogously is to maintain relatedness among both types of entities. The viability of both models is dependent on the assumptions that pedestrian walking behavior is an affective expression in line to each person's motivations and expectations. The conducted experiments made use of real world and simulated scenarios aiming to test the proposed model in a wide variety of circumstances. Results obtained from such experi-

ments indicate an acceptable accuracy in predicting sub-crowd motivations as well as a high correlation between pedestrian and sub-crowd emotions, confirming that the proposed sub-crowd emotions indeed serve as a meaningful abstraction of pedestrian emotions. The experiments conducted here were limited to the environment of a train station; therefore, in future efforts, it is essential to evaluate the method's performance under different types of crowded environments. Another critical aspect to address in the future, is the acquisition of well-validated pedestrian emotion annotations to further support the validity of this method.



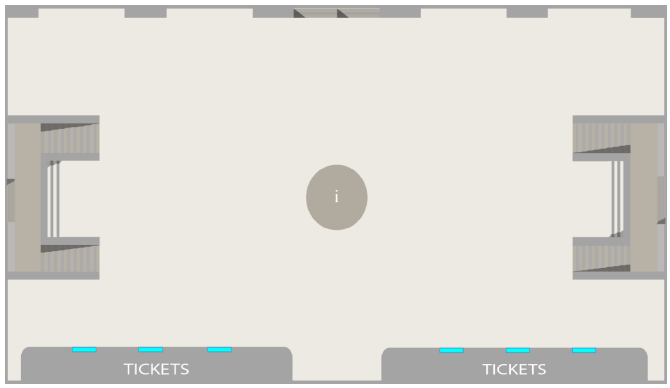
(a)



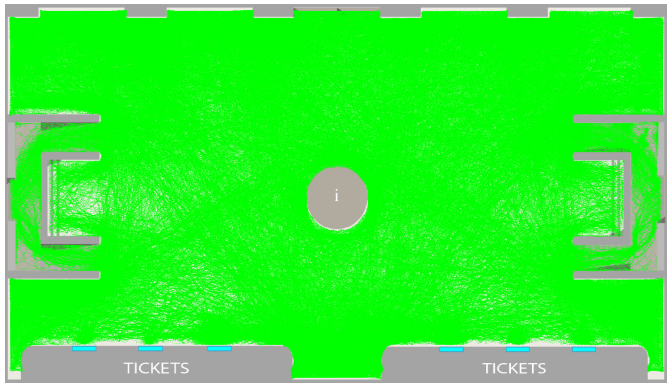
(b)

(c)

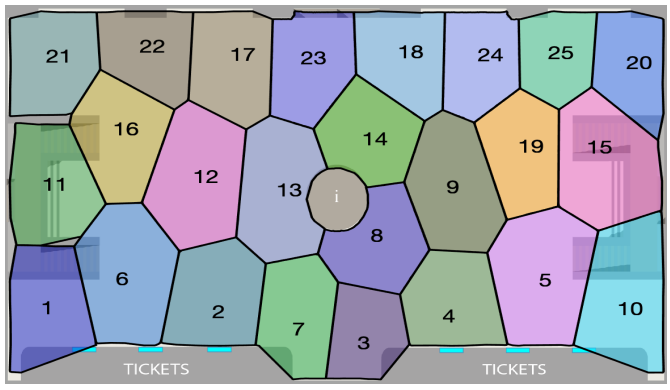
Figure 8.3: Grand central station dataset with (a) the observed environment, (b) training set with annotations colored in green dots, and (c) the environment divided into subregions, displayed with random colors and enumerated by their index.



(a)



(b)



(c)

Figure 8.4: SC-Station dataset presenting (a) the simulated environment, (b) annotated trajectories of the training set in green lines, and (c) the environment divided into subregions colored with random colors and enumerated by their index.

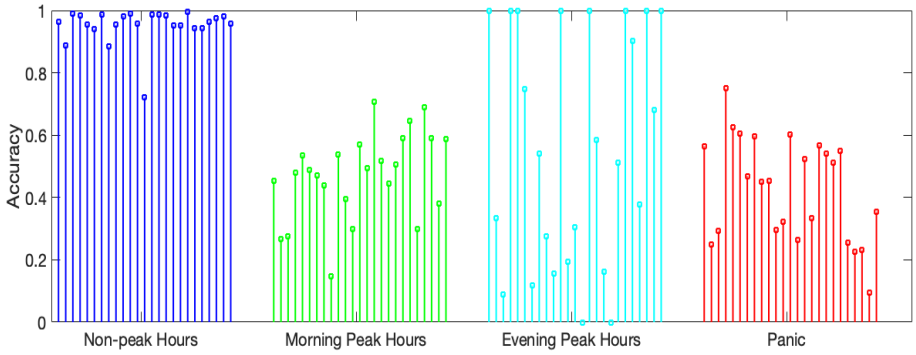


Figure 8.5: Crowd motivation estimation for each subregion in every test set of the SC-Station dataset.

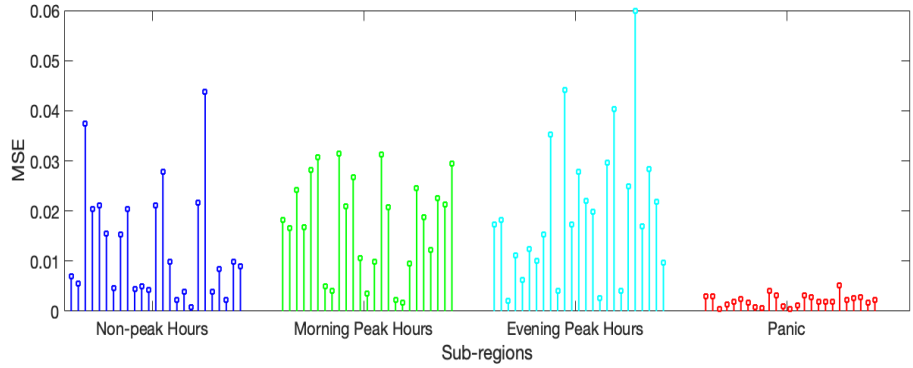


Figure 8.6: Crowd emotion estimation MSE for each subregion in every test set of the SC-Station dataset.

Chapter 9

Conclusions, Limitations and Future Work

This chapter provides a summary of the work presented in the previous chapters and how it addressed our research questions. Next, we outline the limitations encountered throughout this research. Finally we describe potential aspects to be studied in the future.

9.1 Conclusions

The goal of this research was centered on investigating the relationship between visually observable behavior of pedestrians in crowded environments and their emotional state. With this in mind, we devised a series of research questions posed in chapter 1, and the remaining chapters proposed answers to these questions. Our main research question stated:

> **Main Research Question:**
> *How to design a model capable to infer emotions of people in crowded environments?*

In the initial stage of our research [129] we were able to identify previous contributions where the inference of emotions was addressed either to the individual or collective level, but not both. Each way of describing a crowd (i.e., as a whole or by single pedestrians) poses different challenges and advantages

depending on the context. Taking on a dualistic view of the crowd, we divided our main research question into two sub-questions, one to address the inference of emotion of individual pedestrians, and another to provide an emotion estimation of the crowd as a whole.

Research Question 1.1:
How to model the individual pedestrians in a way that allows to associate their walking movement to their underlying emotions?

In addressing the inference of pedestrian emotions, our work in [128] and chapter 5 presented the first iteration of the pedestrian model for the estimation of individual emotions of pedestrians under complex, crowded environments. In comparison to the closest work found in the existing literature [5], our approach provided significant improvements: (1) accounted for scenarios with multiple origin and destination points, (2) Introduced the concepts of motivations and expectations as the building blocks to estimate emotions. And (3) presented the idea of representing behaviors in higher abstractions using words and vocabularies, which helps to reduce data sparsity. The conducted experiments in that chapter confirmed the viability of our model to estimate pedestrian motivations, model their behavior, and make use of this to produce an estimation of their emotional state. The pedestrian model was further refined in our work [131] and in chapter 6 where an improved version included the addition of expectations specific to individual pedestrians to infer their emotions. The prediction of pedestrian's motivation was addressed using direction fields generated for each POI. The emotional state was derived from the difference between expected and actual observed behaviors. A hypothesis of expectation for pedestrians in a train station was proposed and employed to generate emotional state annotations. The results of the proposed model indicated a significant improvement for predicting motivations (destinations) over previous works, and to efficiently estimate individual emotions based on the proposed hypothesis for expectations.

The innovative contributions of the pedestrian model include: (a) a data-driven model based on hierarchical Bayesian networks that includes multiple levels of abstraction to account for behavioral and psychological factors, and is capable to generalize to different contexts in a supervised way; (b) the introduction of the distance-to-motivation (DTM) measurement which helps to form an association between observable behavior and emotions with strong foundation in psychological principles; (c) an emotion annotation scheme for automatic labeling of pedestrian trajectories based on learned motivations and expectations;

and (d) A method to infer emotions in a continuum valence axis, extending from abnormality detection or simple behavior classification, a recurring theme in the crowd emotion models discussed in our literature review [14] [91] [6].

Research Question 1.2:
How to produce an inference of the collective emotion of pedestrians in a way that is consistent with their individual emotions?

Turning our attention to estimate the collective emotion of pedestrians, in [130] and chapter 7 we presented the first iteration of the crowd model aiming to explore a suitable approach to describe the dynamics of a crowd as a whole, under realistic crowded environments. Behaviors were modeled by changes in density distribution in the crowd rather than by observing single trajectories of pedestrians. This approach was advantageous when crowd density was high as in such circumstances, people counting methods perform better than pedestrian tracking techniques. Also, seen the crowd as a whole provided a more comprehensive understanding of the crowd's dynamics. The conducted experiments supported the viability of achieving a macroscopic view of the crowd yielding consistently reliable predictions of its future behavior and identification of motivation. In particular, the crowd model performed with high accuracy to identify different behaviors in crowds even when the predominant behavior was exhibited by as low as only half of the pedestrians in the crowd. Later in chapter 8 the model was extended to account for crowds within crowds and provided a formal definition of behavior, motivation, expectation, and emotion at the sub-crowd level that was analogous to the pedestrian model. The intention to model both the pedestrian and the crowd analogously was to maintain relatedness among both types of entities. The viability of both models was dependent on the assumptions that pedestrian walking behavior was an affective expression in line to each person's motivations and expectations. The conducted experiments made use of real world and simulated scenarios aiming to test the proposed model in a wide variety of circumstances. Results obtained from such experiments indicated an acceptable accuracy in predicting sub-crowd motivations as well as a high correlation between pedestrian and sub-crowd emotions, confirming that the proposed sub-crowd emotions indeed serve as a meaningful abstraction of pedestrian emotions.

Innovations introduced with the crowd model include: (a) a data-driven model based on hierarchical Bayesian networks capable to generalize to ambulatory crowds in multiple contexts, and that expands the concepts of motivation, expectation and emotions from the individual pedestrian to a collective level in

a consistent and equivalent way for both the pedestrian and the crowd; (b) a method to describe crowds and sub-crowds based on spatial-temporal interactions learned from partial observation of pedestrian trajectories; and (c) an approach to infer the collective emotion of crowds and sub-crowds measured in a continuum valence axis, in a way that is representative and consistent with the emotions experienced by the pedestrians member of the crowd.

9.2 Limitations

Undoubtedly we are faced with several limitations when trying to infer the emotion of pedestrians solely from visually observed walking trajectories. The first and foremost limitation is that the models presented in this thesis apply only to crowds with a predominantly ambulatory behavior, as is the case of casual crowds, queues, acquisitive, mobs, riots, and panic crowds. On the other hand, our models are not suitable for passive crowds like audiences, spectator crowds, and information-seeking groups. Secondly, as we employ surveillance cameras to observe the crowd due to their ubiquitousness in public spaces, we depend on the performance of crowd counting and pedestrian detection & tracking algorithms. This entails that in the presence of highly dense crowds, pedestrian detection & tracking algorithms will deliver fragmented trajectories, tampering the performance of the pedestrian model. Another relevant limitation concerns the absence of context-awareness in our method, i.e., the association of learned behavior to an emotion label is dependent on the context of the situation. To illustrate this point, consider a group of kids running in a park, where their behavior may be associated with valence in the positive spectrum while the same behavior displayed by adults in a train station may correspond to a panic situation where the associated valence lies on the negative spectrum. For this reason, models of behaviors learned unsupervisedly need to be empirically labeled by a human operator. Finally, the emotion inferences produced by both the pedestrian and crowd models are approximations that include only the valence dimension and account merely for the stimuli related to the pedestrian's intention to reach a destination in the environment. Therefore, our approach is unable to draw any assumption about the emotions evoked in pedestrians from internal stimuli or factors unrelated to their desire to reach a destination.

9.3 Practical Applications

Appreciating the frequent occurrence of crowds and the important role of emotions in understanding the dynamics of crowds, the pedestrian and crowd models developed throughout this thesis have a wide range of practical applications. In the general sense, as our models learn to associate observed behaviors with manual labels of emotions, the predictive capabilities of what emotions will arise from the exhibited behavior of crowds can result useful in several areas concerning pedestrian gatherings. Taking into consideration the capabilities and limitations of our work, as follow we list some potential practical applications where we consider our models can prove to be beneficial.

Crowd Management: this area concerns the development and implementation of best practices to ensure the security and safeness of large gatherings of people, such as those seeing in sport events and concerts. Aiming to reduce the intervention of human operators, current implementations of surveillance systems allow the automation of certain tasks like abnormal events detection. In this context, where a macroscopic view of the crowd is more relevant, our crowd model can help in providing early warnings when minor levels of negative emotions are detected before an actual incident takes place, given that our method infers emotions in a continuous granularity rather than binary (normal/abnormal) states. For instance, using our method to learn about the expected behavior of a queuing sub-crowd in the environment, we can signal when waiting times are exceeding acceptable thresholds, helping to deescalate a potentially violent situation.

Urban planning: this process addresses the design and development of public spaces for the use of general population. The use of crowd simulations to understand how pedestrians move and behave in a particular environment is already a common practice, however these methods are intended to inform on the efficiency of space utilization without considering psychological factors only observable in real-life scenarios. Implementing our method in existing infrastructures can allow to gain insights into recurrent patterns of how people's emotions are influenced by the infrastructure or environmental factors. As an example, let's consider a train station where the new placement of ticket windows is convenient to find for tourists but causes subtle inconvenience for a large number of passengers walking towards some exit doors; such situation will be identified by our models as it will impact the expected flow of pedestrians, resulting in measurable variations of the inferred emotions.

Smart Environments: At is most basic form, it refers to infrastructures equipped with automated processes to control certain operations as lighting and ventilation. A common case of this approach is the use of motion sensors to turn lights on only when people are present, hence reducing electricity consumption. In a more complex example, we can envision a smart building supplied with a cognitive dynamic system, i.e., a system that builds up rules of behavior over time through learning from continuous experiential interactions with the environment. In this case, our model can provide an inference of the emotion of the crowd in real time, enabling the system to make decisions on how to control pedestrian flow by opening/closing doors, dynamic signal display or other means.

9.4 Future Work

In the pursue of reliable and robust models capable to infer emotions of people in crowded environments, several gaps in the literature remain open for exploration.

From the limited amount of research devoted to the estimation of emotions in crowds, a great majority focuses only on the types of scenarios directly associated with the negative spectrum of emotions, as is the case of panic, evacuations and riots [79]. The unbalanced interest in these scenarios is well justified given that safety is a major concern in crowded environments; its also advantageous that a behavior displayed in such undesired conditions is distinctively different from its relative normal behavior. The models presented in this thesis focused on a more generalized spectrum of circumstances by providing the means to learn associations between observed behaviors and their corresponding emotions along a valence-axis, ranging from negative to neutral and positive affective states. An aspect for potential improvement of this model is the inclusion of the arousal-axis, completing the factors considered by the family of dimensional theories of emotions [109] over which we based our research. Further developing models that fully adopt the selected theory of emotion would increase their robustness and reliability, encouraging real-world implementations.

The experiments conducted in the previous chapters employed real-world and simulated datasets corresponding to train stations and shopping malls. These environments were deliberately chosen because they offered the benefit of a large open space where the area of interest could be observed with a single surveillance camera; However, this puts in doubt how well the proposed

models can generalize to different conditions decisions since it limits the category of observed behaviors. For instance, the dynamics of pedestrians in a train station is significantly distant to that in school playgrounds or museums. Hence, another point of improvement lies in conducting additional experiments to test the performance of our proposed models in more diverse situations.

Another reason for limiting the environments explored in the conducted experiments was the lack publicly available datasets that provide well validated annotations of emotional states. A survey conducted in our previous work [129] evidenced a relative abundance of crowd-related datasets providing annotations of trajectories, density levels, and categorical behaviors, neglecting the affective aspects of the crowd. An example of the scarce datasets providing annotations related to emotions is Rabiee's contribution [104] which captures a crowd in a single environment and produce annotations of panic, fight, congestion, obstacle, and neutral behaviors. This shortage of options is justified considering the current interest in crowd analysis comes from the field of computer vision where the tasks of crowd density estimation, people detection and tracking, group actions, and collectiveness & cohesiveness occupy a higher concern. Additional constrains to designing datasets richer in emotional diversity come from ethical concerns in subjecting people to potentially harmful situations. One way to overcome ethical concerns is to utilize existing footage of past events where crowds were observed in the conditions we aim to study. one more option is to employ simulation tools, at the cost of compromising the naturalness and credibility of the produced data. The absence of available datasets and common benchmarks further impedes the comparison and improvement of methods intended for emotion estimation in crowds.

Based on our current understanding of the decision-making process [32] [11] [120] guiding our actions, we identify a cyclic relationship between emotions and behavioral responses. The research conducted in this thesis focused on inferring emotions from observable behavior, leaving the reverse case of this relationship uninvestigated. Having emotion estimation models reach a level of maturity where affective states can be inferred with a higher degree of confidence, it opens the possibility to explore how emotions can be used to predict the behavior of single pedestrians and the crowd.

Bibliography

[1] ADOLPHS, R., DAMASIO, H., TRANEL, D., AND DAMASIO, A. R. Cortical systems for the recognition of emotion in facial expressions. *The Journal of neuroscience : the official journal of the Society for Neuroscience 16*, 23 (1996), 7678–7687.

[2] ADRIAN, J., AMOS, M., BARATCHI, M., BEERMANN, M., BODE, N., BOLTES, M., CORBETTA, A., DEZECACHE, G., DRURY, J., FU, Z., ET AL. A glossary for research on human crowd dynamics. *Collective Dynamics 4*, A19 (2019), 1–13.

[3] ALI, S., AND SHAH, M. Floor fields for tracking in high density crowd scenes. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (2008), pp. 1–14.

[4] ANTONINI, G., MARTINEZ, S. V., BIERLAIRE, M., AND THIRAN, J. P. Behavioral priors for detection and tracking of pedestrians in video sequences. *International Journal of Computer Vision 69*, 2 (2006), 159–180.

[5] BAIG, M. B., BARAKOVA, E., MARCENARO, L., REGAZZONI, C., AND RAUTERBERG, M. Bio-inspired probabilistic model for crowd emotion detection. In *Proceedings of the International Joint Conference on Neural Networks* (2014), pp. 3966–3973.

[6] BAIG, M. W., BAIG, M. S., BASTANI, V., BARAKOVA, E. I., MARCENARO, L., REGAZZONI, C. S., AND RAUTERBERG, M. Perception of emotions

from crowd dynamics. In *2015 IEEE International Conference on Digital Signal Processing (DSP)* (2015), IEEE, pp. 703–707.

[7] BARAKOVA, E. I., GORBUNOV, R., AND RAUTERBERG, M. Automatic Interpretation of Affective Facial Expressions in the Context of Interpersonal Interaction. *IEEE Transactions on Human-Machine Systems 45*, 4 (2015), 409–418.

[8] BARLIYA, A., OMLOR, L., GIESE, M. A., BERTHOZ, A., AND FLASH, T. Expression of emotion in the kinematics of locomotion. *Experimental brain research 225*, 2 (2013), 159–176.

[9] BATTY, M., AXHAUSEN, K. W., GIANNOTTI, F., POZDNOUKHOV, A., BAZZANI, A., WACHOWICZ, M., OUZOUNIS, G., AND PORTUGALI, Y. Smart cities of the future. *European Physical Journal: Special Topics* (2012).

[10] BAVEYE, Y. B., BETTINELLI, J.-N., DELLANDRÉA, E., CHEN, L., AND CHAMARET, C. A large video data base for computational models of induced emotion. In *Proceedings - 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013* (2013), pp. 13–18.

[11] BECHARA, A., AND DAMASIO, A. R. The somatic marker hypothesis: A neural theory of economic decision. *Games and economic behavior 52*, 2 (2005), 336–372.

[12] BEIK, W. The violence of the french crowd from charivari to revolution. *Past and Present 197*, 1 (2007), 75–110.

[13] BERLONGHI, A. E. Understanding and planning for different spectator crowds. *Safety Science 18*, 4 (1995), 239–247.

[14] BOSSE, T., DUELL, R., MEMON, Z. A., TREUR, J., AND VAN DER WAL, C. N. Agent-Based Modeling of Emotion Contagion in Groups. *Cognitive Computation 7*, 1 (2014), 111–136.

[15] BOSSE, T., HOOGENDOORN, M., KLEIN, M., SHARPANSKYKH, A., TREUR, J., VAN DER WAL, C. N., AND VAN WISSEN, A. Agent-based modelling of social emotional decision making in emergency situations. In *Co-Evolution of Intelligent Socio-Technical Systems*. Springer, 2013, pp. 79–117.

[16] BOSSE, T., HOOGENDOORN, M., KLEIN, M. C., TREUR, J., AND VAN DER WAL, C. N. Agent-based analysis of patterns in crowd behaviour involving contagion of mental states. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems* (2011), Springer, pp. 566–577.

[17] BOSSE, T., HOOGENDOORN, M., KLEIN, M. C. A., TREUR, J., VAN DER WAL, C. N., AND VAN WISSEN, A. Modelling collective decision making in groups and crowds: Integrating social contagion and interacting emotions, beliefs and intentions. *Autonomous Agents and Multi-Agent Systems 27*, 1 (2013), 52–84.

[18] BOUWMANS, T., SILVA, C., MARGHES, C., ZITOUNI, M. S., BHASKAR, H., AND FRELICOT, C. On the role and the importance of features for background modeling and foreground detection. *Computer Science Review 28* (2018), 26–91.

[19] BROWN, C., AND LEWIS, E. L. Protesting the invasion of cambodia: A case study of crowd behavior and demonstration leadership. *Polity 30*, 4 (1998), 645–665.

[20] BURNES, B., AND COOKE, B. Kurt Lewin's field theory: A review and re-evaluation. *International Journal of Management Reviews 15*, 4 (2013), 408–425.

[21] BUSSO, C., BULUT, M., LEE, C.-C., KAZEMZADEH, A., MOWER, E., KIM, S., CHANG, J. N., LEE, S., AND NARAYANAN, S. S. Iemocap: Interactive emotional dyadic motion capture database. *Language resources and evaluation 42*, 4 (2008), 335.

[22] CAMBIAGHI, M., AND SACCHETTI, B. Ivan petrovich pavlov (1849–1936). *Journal of neurology 262*, 6 (2015), 1599–1600.

[23] CAO, H., COOPER, D., KEUTMANN, M., GUR, R. E., NENKOVA, A., AND VERMA, R. CREMA-D: Crowd-sourced emotional multimodal actors dataset. *IEEE Transactions on Affective Computing 5*, 4 (2014), 377–390.

[24] CHALLENGER, R., CLEGG, C. W., AND ROBINSON, M. *Understanding crowd behaviours: Practical guidance and lessons identified*. TSO, 2010.

[25] CHALLENGER, R., CLEGG, C. W., AND ROBINSON, M. A. Understanding crowd behaviours: Supporting evidence. *Understanding Crowd Behaviours* (2009), 1–326.

[26] CHANDRAN, A. K., POH, L. A., AND VADAKKEPAT, P. Identifying social groups in pedestrian crowd videos. In *International Conference on Advances in Pattern Recognition* (2015), IEEE, pp. 1–6.

[27] CHENG, Z., QIN, L., HUANG, Q., YAN, S., AND TIAN, Q. Recognizing human group action by layered model with multiple cues. *Neurocomputing 136* (2014), 124–135.

[28] CHIAPPINO, S., MORERIO, P., MARCENARO, L., FUIANO, E., REPETTO, G., AND REGAZZONI, C. S. A multi-sensor cognitive approach for active security monitoring of abnormal overcrowding situations. In *International Conference on Information Fusion* (2012), IEEE, pp. 2215–2222.

[29] CHIAPPINO, S., MORERIO, P., MARCENARO, L., AND REGAZZONI, C. S. Bio-inspired relevant interaction modelling in cognitive crowd management. *Journal of Ambient Intelligence and Humanized Computing 6*, 2 (2015), 171–192.

[30] CHICKERUR, S., AND JOSHI, K. 3D face model dataset: Automatic detection of facial expressions and emotions for educational environments. *British Journal of Educational Technology 46*, 5 (2015), 1028–1037.

[31] CHO, S.-Y., CHOW, T. W., AND LEUNG, C.-T. A neural-based crowd estimation by hybrid global learning algorithm. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 29*, 4 (1999), 535–541.

[32] CHOHRA, A., AND MADANI, K. Biological regulation and psychological mechanisms models of adaptive decision-making behaviors: drives, emotions, and personality. In *Transactions on Computational Collective Intelligence XXIX*. Springer, 2018, pp. 69–83.

[33] COOK, D., AND DAS, S. K. *Smart environments: technology, protocols, and applications*, vol. 43. John Wiley & Sons, 2004.

[34] CORBETTA, A., MENKOVSKI, V., AND TOSCHI, F. Weakly supervised training of deep convolutional neural networks for overhead pedestrian localization in depth fields. In *International Conference on Advanced Video and Signal Based Surveillance (AVSS)* (2017), IEEE, pp. 1–6.

[35] CORTES, C., AND VAPNIK, V. Support-vector networks. *Machine learning 20*, 3 (1995), 273–297.

[36] CRAGGS, R., AND WOOD, M. A two dimensional annotation scheme for emotion in dialogue. In *Proceedings of AAAI spring symposium: exploring attitude and affect in text* (2004), vol. 102.

[37] CUCCHIARA, R., GRANA, C., PICCARDI, M., AND PRATI, A. Detecting moving objects, ghosts, and shadows in video streams. *IEEE transactions on pattern analysis and machine intelligence 25*, 10 (2003), 1337–1342.

[38] DALAL, N., AND TRIGGS, B. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)* (2005), vol. 1, IEEE, pp. 886–893.

[39] DAMASIO, A. R. The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences 351*, 1346 (1996), 1413–20.

[40] DAS, D., AND CHAKRABARTY, A. Emotion Recognition from Face Dataset Using Deep Neural Nets. *IET Computer Vision 10*, 6 (2016), 81–100.

[41] DE GELDER, B., VAN DEN STOCK, J., MEEREN, H. K., SINKE, C. B., KRET, M. E., AND TAMIETTO, M. Standing up for the body. recent progress in uncovering the networks involved in the perception of bodies and bodily expressions. *Neuroscience & Biobehavioral Reviews 34*, 4 (2010), 513–527.

[42] DICKIE, J. Major crowd catastrophes. *Safety science 18*, 4 (1995), 309–320.

[43] DOUGLAS-COWIE, E., COWIE, R., SNEDDON, I., COX, C., LOWRY, O., MCRORIE, M., MARTIN, J.-C., DEVILLERS, L., ABRILIAN, S., BATLINER, A., ET AL. The humaine database: Addressing the collection and annotation of naturalistic and induced emotional data. In *International conference on affective computing and intelligent interaction* (2007), Springer, pp. 488–500.

[44] DRURY, J., AND COCKING, C. *The mass psychology of disasters and emergency evacuations: A research report and implications for practice*. University of Sussex Brighton, 2007.

[45] DUIVES, D. C., DAAMEN, W., AND HOOGENDOORN, S. P. State-of-the-art crowd motion simulation models. *Transportation Research Part C: Emerging Technologies 37* (2014), 193–209.

[46] EKMAN, P. An argument for basic emotions. *Cognition & emotion 6*, 3-4 (1992), 169–200.

[47] EKMAN, P. Facial expression and emotion. *American psychologist 48*, 4 (1993), 384.

[48] EKMAN, P., FRIESEN, W. V., O'SULLIVAN, M., CHAN, A., DIACOYANNI-TARLATZIS, I., HEIDER, K., KRAUSE, R., LECOMPTE, W. A., PITCAIRN, T., RICCI-BITTI, P. E., ET AL. Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of personality and social psychology 53*, 4 (1987), 712.

[49] EKMAN, R. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, USA, 1997.

[50] ENDO, N., ENDO, K., HASHIMOTO, K., KOJIMA, T., IIDA, F., AND TAKANISHI, A. Integration of emotion expression and visual tracking locomotion based on vestibulo-ocular reflex. In *19th International Symposium in Robot and Human Interactive Communication* (2010), IEEE, pp. 558–563.

[51] FERRYMAN, J., AND ELLIS, A. L. Performance evaluation of crowd image analysis using the PETS2009 dataset. *Pattern Recognition Letters 44* (2014), 3–15.

[52] FINUCANE, M. L., ALHAKAMI, A., SLOVIC, P., AND JOHNSON, S. M. The affect heuristic in judgments of risks and benefits. *Journal of behavioral decision making 13*, 1 (2000), 1–17.

[53] FOXALL, G. R. Intentional behaviorism. *Behavior and Philosophy* (2007), 1–55.

[54] FREUD, S. Group psychology and the analysis of the ego. In *The Standard Edition of the Complete Psychological Works of Sigmund Freud, Volume XVIII (1920-1922): Beyond the Pleasure Principle, Group Psychology and Other Works*. 1955, pp. 65–144.

[55] FRUIN, J. Designing for Pedestrians: A Level-of-Service Concept. New York Metropolitan Association of Urban Designers and Environmental Planners. *Highway Research Record*, 355 (1971).

[56] FRUIN, J. J. Designing for Pedestrians: a Level-of-Service Concept. *Highway Research Record 377* (1971), 1–15.

[57] GOODFELLOW, I. J., ERHAN, D., LUC CARRIER, P., COURVILLE, A., MIRZA, M., HAMNER, B., CUKIERSKI, W., TANG, Y., THALER, D., LEE, D. H., ZHOU, Y., RAMAIAH, C., FENG, F., LI, R., WANG, X., ATHANASAKIS, D., SHAWE-TAYLOR, J., MILAKOV, M., PARK, J., IONESCU, R., POPESCU, M., GROZEA, C., BERGSTRA, J., XIE, J., ROMASZKO, L., XU, B., CHUANG, Z., AND BENGIO, Y. Challenges in representation learning: A report on three machine learning contests. *Neural Networks 64* (2015), 59–63.

[58] GRANOVETTER, M. Threshold models of collective behavior. *American journal of sociology 83*, 6 (1978), 1420–1443.

[59] GRANT, J. M., AND FLYNN, P. J. Crowd scene understanding from video: a survey. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 13*, 2 (2017), 1–23.

[60] GROSS, H. S. Looking for Spinoza: Joy, Sorrow, and the Feeling Brain. *The Journal of Nervous and Mental Disease* (2004).

[61] GUNES, H., AND PANTIC, M. Automatic, Dimensional and Continuous Emotion Recognition. *International Journal of Synthetic Emotions 1*, 1 (2010), 68–99.

[62] HASSAN, M. A., MALIK, A. S., NICOLAS, W., FAYE, I., AND NORDIN, N. Reliability of bench-mark datasets for crowd analytic surveillance. *2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings 2015-July* (2015), 1084–1089.

[63] HEALTH, AND EXECUTIVE, S. The Event Safety Guide: A Guide to Health, Safety and Welfare at Music and Similar Events. *Norwich : HSE Books* (1999).

[64] HELBING, D., AND MOLNAR, P. Social force model for pedestrian dynamics. *Physical review E 51*, 5 (1995), 4282.

[65] HELBING, D., AND MOLNAR, P. Social force model for pedestrian dynamics. *Physical review E 51*, 5 (1995), 4282.

[66] HOGG, M. A., AND TINDALE, R. S. Blackwell Handbook of Social Psychology: Group Processes. *Group 124* (2001), 1–696.

[67] HOOGENDOORN, M., TREUR, J., VAN DER WAL, C. N., AND VAN WISSEN, A. Modelling the interplay of emotions, beliefs and intentions within collective decision making based on insights from social neuroscience. *Lecture Notes in Computer Science 6443 LNCS*, PART 1 (2010), 196–206.

[68] HUEBNER, B. Genuinely collective emotions. *European Journal for Philosophy of Science 1*, 1 (2011), 89–118.

[69] KATSIMEROU, C. Crowdsourcing Empathetic Intelligence : The Case of the Annotation of EMMA Database for Emotion and Mood Recognition. *Acm Tist 7*, 4 (2016), 51.

[70] KEMPER, T. D., AND LAZARUS, R. S. Emotion and Adaptation. *Contemporary Sociology 21*, 4 (1992), 522.

[71] KOELSTRA, S., MUHL, C., SOLEYMANI, M., LEE, J.-S., YAZDANI, A., EBRAHIMI, T., PUN, T., NIJHOLT, A., AND PATRAS, I. Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing 3*, 1 (2011), 18–31.

[72] KOHONEN, T. The Self-Organizing Map. *Proceedings of the IEEE 78*, 9 (1990), 1464–1480.

[73] KÖSTER, G., SEITZ, M., TREML, F., HARTMANN, D., AND KLEIN, W. On modelling the influence of group formations in a crowd. *Contemporary Social Science 6*, 3 (2011), 397–414.

[74] LE BON, G. *The crowd: A study of the popular mind*. T. Fisher Unwin, 1897.

[75] LEACH, M. J., BAXTER, R., ROBERTSON, N. M., AND SPARKS, E. P. Detecting social groups in crowded surveillance videos using visual attention. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (2014), pp. 461–467.

[76] LI, B., JIN, Q.-S., AND GUO, H.-Y. Research on Psychological Characteristics of Passengers in Terminal and Some Related Measures. *Procedia - Social and Behavioral Sciences 96*, Cictp (2013), 993–1000.

[77] LI, J., CAI, R., DE RIDDER, H., VERMEEREN, A., AND VAN EGMOND, R. A study on relation between crowd emotional feelings and action tendencies. In *The 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational* (2014), pp. 775–784.

[78] LI, W., ABTAHI, F., TSANGOURI, C., AND ZHU, Z. Towards an "in-the-wild" emotion dataset using a game-based framework. In *Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (2016), IEEE, pp. 1526–1534.

[79] LIU, Z., AND HUANG, P. Study of panic behavior model for crowd on pedestrian bridge in emergent event. *Xitong Fangzhen Xuebao / Journal of System Simulation 24*, 9 (2012), 1950–1953.

[80] LOFLAND, J. *Protest: Studies of collective behaviour and social movements*. Routledge, 2017.

[81] LOWE, R., BARAKOVA, E., BILLING, E., AND BROEKENS, J. Grounding emotions in robots: an introduction to the special issue. *Adaptive Behavior 24*, 5 (2016), 263–266.

[82] LUCEY, P., COHN, J. F., KANADE, T., SARAGIH, J., AMBADAR, Z., AND MATTHEWS, I. The extended cohn-kande dataset (CK+): A complete facial expression dataset for action unit and emotion specified expression. *Cvprw*, July (2010), 94–101.

[83] MANZOOR, A., AND TREUR, J. An agent-based model for integrated emotion regulation and contagion in socially affected decision making. *Biologically Inspired Cognitive Architectures 12* (2015), 105–120.

[84] MCGILL, V., AND WELCH, L. A behaviorist analysis of emotions. *Philosophy of Science 13*, 2 (1946), 100–122.

[85] MCHUGH, J. E., MCDONNELL, R., O'SULLIVAN, C., AND NEWELL, F. N. Perceiving emotion in crowds: The role of dynamic body postures on the perception of emotion in crowded scenes. *Experimental Brain Research 204*, 3 (2010), 361–372.

[86] MCKEOWN, G., VALSTAR, M., COWIE, R., PANTIC, M., AND SCHRODER, M. The semaine database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent. *IEEE transactions on affective computing 3*, 1 (2011), 5–17.

[87] MCPHAIL, C. *The myth of the madding crowd*. Routledge, 2017.

[88] MEHRAN, R., OYAMA, A., AND SHAH, M. Abnormal crowd behavior detection using social force model. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (2009), IEEE, pp. 935–942.

[89] MITCHELL, T. Cmu face images data set. *UCI Machine Learning Repository* (1997).

[90] MOORS, A., ELLSWORTH, P. C., SCHERER, K. R., AND FRIJDA, N. H. Appraisal theories of emotion: State of the art and future development. *Emotion Review 5*, 2 (2013), 119–124.

[91] MOU, W., GUNES, H., AND PATRAS, I. Automatic recognition of emotions and membership in group videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (2016), pp. 27–35.

[92] MOUSAVI, H., MOHAMMADI, S., PERINA, A., CHELLALI, R., AND MURINO, V. Analyzing tracklets for the detection of abnormal crowd behavior. In *Winter Conference on Applications of Computer Vision* (2015), IEEE, pp. 148–155.

[93] MURPHY, K. P., AND RUSSELL, S. *Dynamic bayesian networks: representation, inference and learning*. PhD thesis, University of California, Berkeley Berkeley, CA, 2002.

[94] MUSSE, S. R., AND THALMANN, D. A model of human crowd behavior: Group inter-relationship and collision detection analysis. In *Computer Animation and Simulation*. Springer, 1997, pp. 39–51.

[95] MYERS, D. G. *Social psychology*. McGraw-Hill, 1996.

[96] NASUKAWA, T., AND YI, J. Sentiment analysis: Capturing favorability using natural language processing. In *Proceedings of the 2nd international conference on Knowledge capture* (2003), pp. 70–77.

[97] NGUYEN, N. T., KOWALCZYK, R., AND MERCIK, J. *Transactions on Computational Collective Intelligence XXIII*, vol. 9760. Springer, 2016.

[98] P. WILSON, J., AND GOSIEWSKA, S. Multi-agency gold incident command training for civil emergencies. *Disaster Prevention and Management 23*, 5 (2014), 632–648.

[99] PAK, A., AND PAROUBEK, P. Twitter as a Corpus for Sentiment Analysis and Opinion Mining Microblogging Microblogging = posting small blog entries Platforms. *Analysis* (2010), 1320–1326.

[100] PERUGIA, G., VAN BERKEL, R., DÍAZ-BOLADERAS, M., CATALÀ-MALLOFRÉ, A., RAUTERBERG, M., AND BARAKOVA, E. Understanding engagement in dementia through behavior. the ethographic and laban-inspired coding system of engagement (elicse) and the evidence-based model of engagement-related behavior (emodeb). *Frontiers in Psychology 9*, MAY (5 2018).

[101] PETERMAN, J., CHRISTENSEN, A., GIESE, M., AND PARK, S. Extraction of social information from gait in schizophrenia. *Psychological medicine 44*, 5 (2014), 987–996.

[102] PLACIDI, G., DI GIAMBERARDINO, P., PETRACCA, A., SPEZIALETTI, M., AND IACOVIELLO, D. Classification of Emotional Signals from the DEAP dataset. *Proceedings of the 4th International Congress on Neurotechnology, Electronics and Informatics*, Neurotechnix (2016), 15–21.

[103] QIN, Z., AND SHELTON, C. R. Social grouping for multi-target tracking and head pose estimation in video. *IEEE transactions on pattern analysis and machine intelligence 38*, 10 (2015), 2082–2095.

[104] RABIEE, H., HADDADNIA, J., MOUSAVI, H., KALANTARZADEH, M., NABI, M., AND MURINO, V. Novel dataset for fine-grained abnormal behavior understanding in crowd. In *13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)* (2016), IEEE, pp. 95–101.

[105] RAHMALAN, H., NIXON, M. S., AND CARTER, J. N. On crowd density estimation for surveillance. IET.

[106] REICHER, S. 'The Crowd' century: Reconciling practical success with theoretical failure. *British Journal of Social Psychology 35*, 4 (1996), 535–553.

[107] REICHER, S. The psychology of crowd dynamics. *Blackwell handbook of social psychology: Group processes* (2001), 182–208.

[108] REICHER, S. Crowd Psychology. *Encyclopedia of Human Behavior: Second Edition* (2012).

[109] RUSSELL, J. A. Core affect and the psychological construction of emotion. *Psychological review 110*, 1 (2003), 145–72.

[110] RUSSELL, J. A., AND MEHRABIAN, A. Evidence for a three-factor theory of emotions. *Journal of Research in Personality* (1977).

[111] SÁNCHEZ, J., PERRONNIN, F., MENSINK, T., AND VERBEEK, J. Image classification with the fisher vector: Theory and practice. *International Journal of Computer Vision 105*, 3 (2013), 222–245.

[112] SARIYANIDI, E., DAGLI, V., TEK, S. C., TUNÇ, B., AND GÖKMEN, M. Local Zernike Moments: A new representation for face recognition. In *Proceedings - International Conference on Image Processing, ICIP* (2012), pp. 585–588.

[113] SCHERER, K. R. Appraisal Theory. *Handbook of Cognition and Emotion* (2005).

[114] SCHERER, K. R. On the rationality of emotions: or, When are emotions rational? *Social Science Information 50*, 3-4 (2011), 330–350.

[115] SEER, S., RUDLOFF, C., MATYUS, T., AND BRÄNDLE, N. Validating social force based models with comprehensive real world motion data. *Transportation Research Procedia 2*, 0 (2014), 724–732.

[116] SETHI, R. J. Towards defining groups and crowds in video using the atomic group actions dataset. *Proceedings - International Conference on Image Processing, ICIP 2015-Decem* (2015), 2925–2929.

[117] SETTI, F., CONIGLIARO, D., ROTA, P., BASSETTI, C., CONCI, N., SEBE, N., AND CRISTANI, M. The s-hock dataset: A new benchmark for spectator crowd analysis. *Computer Vision and Image Understanding 159* (2017), 47–58.

[118] SHARMA, A. Crowd-behavior prediction using subjective factor based multi-agent system. In *International conference on systems, man and cybernetics.* (2000), vol. 1, IEEE, pp. 298–300.

[119] SHARPANSKYKH, A., AND TREUR, J. An adaptive agent model for affective social decision making. *Biologically Inspired Cognitive Architectures 5* (2013), 72–81.

[120] SHARPANSKYKH, A., AND ZIA, K. Emotional decision making in large crowds. *Advances in Intelligent and Soft Computing 155 AISC* (2012), 191–200.

[121] SHUAIBU, A. N., MALIK, A. S., AND FAYE, I. Behavior representation in visual crowd scenes using space-time features. In *2016 6th International Conference on Intelligent and Advanced Systems (ICIAS)* (2016), IEEE, pp. 1–6.

[122] SMITH, T. W. *The Book of Human Emotions: From Ambiguphobia to Umpty – 154 Words from Around the World for How We Feel*. Little, Brown, 2016.

[123] SOLERA, F., CALDERARA, S., AND CUCCHIARA, R. Learning to identify leaders in crowd. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (2015), pp. 43–48.

[124] SPRENGELMEYER, R., RAUSCH, M., EYSEL, U. T., AND PRZUNTEK, H. Neural structures associated with recognition of facial expressions of basic emotions. *Proceedings of the Royal Society of London. Series B: Biological Sciences 265*, 1409 (1998), 1927–1931.

[125] STAUFFER, C., AND GRIMSON, W. E. L. Adaptive background mixture models for real-time tracking. In *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (1999), vol. 2, IEEE, pp. 246–252.

[126] TILLY, C. From Mobilization to Revolution Addison-Wesley. *Reading, Mass* (1978).

[127] TURNER, J. C. Towards a cognitive redefinition of the social group. *Social identity and intergroup relations* (1982), 15–40.

[128] URIZAR, O. J., BAIG, M. S., BARAKOVA, E. I., REGAZZONI, C. S., MARCENARO, L., AND RAUTERBERG, M. A hierarchical bayesian model for crowd emotions. *Frontiers in computational neuroscience 10* (2016), 63.

[129] URIZAR, O. J., BARAKOVA, E. I., MARCENARO, L., REGAZZONI, C. S., AND RAUTERBERG, M. Emotion estimation in crowds: a survey. In *Proceedings of International Conference of Pattern Recognition Systems* (2017), IET, pp. 149–154.

[130] URIZAR, O. J., BARAKOVA, E. I., REGAZZONI, C. S., AND RAUTERBERG, M. Modeling crowds as single-minded entities. In *2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)* (2016), IEEE, pp. 1186–1191.

[131] URIZAR, O. J., MARCENARO, L., REGAZZONI, C. S., BARAKOVA, E. I., AND RAUTERBERG, M. Emotion estimation in crowds: the interplay of motivations and expectations in individual emotions. In *European Signal Processing Conference (EUSIPCO)* (2018), IEEE, pp. 1092–1096.

[132] VAN DEN STOCK, J., RIGHART, R., AND DE GELDER, B. Body expressions influence recognition of emotions in the face and voice. *Emotion 7*, 3 (2007), 487.

[133] VAN HOOREN, B., GOUDSMIT, J., RESTREPO, J., AND VOS, S. Real-time feedback by wearables in running: Current approaches, challenges and suggestions for improvements. *Journal of Sports Sciences 38*, 2 (2020), 214–230.

[134] VON SCHEVE, C., ISMER, S., SCHEVE, C. V., AND ISMER, S. Towards a Theory of Collective Emotions. *Emotion Review 5*, 4 (2013), 406–413.

[135] WANG, B., YE, M., LI, X., AND ZHAO, F. Abnormal crowd behavior detection using size-adapted spatio-temporal features. *International Journal of Control, Automation and Systems 9*, 5 (2011), 905.

[136] WOOD, I., MCCRAE, J. P., ANDRYUSHECHKIN, V., AND BUITELAAR, P. A comparison of emotion annotation schemes and a new annotated data set. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)* (2018).

[137] YI, S., LI, H., AND WANG, X. Understanding pedestrian behaviors from stationary crowd groups. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 07-12-June* (2015), 3488–3496.

[138] YOGAMEENA, B., AND PRIYA, K. S. Synoptic video based human crowd behavior analysis for forensic video surveillance. In *International Conference on Advances in Pattern Recognition (ICAPR)* (2015), IEEE, pp. 1–6.

[139] YOUNG, A. W., PERRETT, D., CALDER, A., SPRENGELMEYER, R., AND EKMAN, P. Facial expressions of emotion: Stimuli and tests (FEEST). *Thames Valley Test Company* (2004).

[140] ZEITZ, K. M., TAN, H. M., GRIEF, M., COUNS, P. C., ZEITZ, C. J., AND STREET, F. Crowd Behavior at Mass Gatherings : A Literature Review. *Prehosp Disaster Med 24*, 1 (2009), 32–8.

[141] ZHAN, B., MONEKOSSO, D. N., REMAGNINO, P., VELASTIN, S. A., AND XU, L.-Q. Crowd analysis: a survey. *Machine Vision and Applications 19*, 5-6 (2008), 345–357.

[142] ZHANG, C., KANG, K., LI, H., WANG, X., XIE, R., AND YANG, X. Data-Driven Crowd Understanding: A Baseline for a Large-Scale Crowd Dataset. *IEEE Transactions on Multimedia 18*, 6 (2016), 1048–1061.

[143] ZHANG, J., KLINGSCH, W., SCHADSCHNEIDER, A., AND SEYFRIED, A. Ordering in bidirectional pedestrian flows and its influence on the fundamental diagram. *Journal of Statistical Mechanics: Theory and Experiment 2012*, 02 (2012), P02002.

[144] ZHOU, B., WANG, X., AND TANG, X. Understanding collective crowd behaviors: Learning a Mixture model of Dynamic pedestrian-Agents. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2012), 2871–2878.

[145] ZHOU, S., CHEN, D., CAI, W., LUO, L., LOW, M. Y. H., TIAN, F., TAY, V. S.-H., ONG, D. W. S., AND HAMILTON, B. D. Crowd modeling and simulation technologies. *ACM Transactions on Modeling and Computer Simulation 20*, 4 (2010), 1–35.

[146] ZITOUNI, M. S., BHASKAR, H., DIAS, J., AND AL-MUALLA, M. E. Advances and trends in visual crowd analysis: A systematic survey and evaluation of crowd modelling techniques. *Neurocomputing 186* (2016), 139–159.

[147] ZITOUNI, M. S., SLUZEK, A., AND BHASKAR, H. Visual analysis of socio-cognitive crowd behaviors for surveillance: A survey and categorization of trends and methods. *Engineering Applications of Artificial Intelligence 82* (jun 2019), 294–312.

[148] ZWENG, A., AND KAMPEL, M. Unexpected human behavior recognition in image sequences using multiple features. In *International Conference on Pattern Recognition* (2010), IEEE, pp. 368–371.

Biography

Oscar Ricardo Juarez Urizar, originally from Guatemala, obtained his double Ph.D. degree from the Design Intelligence group at the Technical University of Eindhoven (Tu/e) and from the Signal Processing group at the University of Genoa (UNIGE). In 2007, he was awarded a full scholarship by the Ministry of Foreign Affairs (MOFA) from the Republic of China (ROC) to learn Mandarin and pursue a bachelor's degree, concluding successfully in 2012 after receiving a BS in Computer Science from National Chiao Tung University (NCTU). He was then a recipient of the NCTU International Scholarship and obtained his MSc in Electrical Engineering and Computer Science in 2014, also from National Chiao Tung University. In 2015 he was selected to join the Erasmus Mundus Joint Doctorate program in Interactive and Cognitive Environments (EMJD-ICE) hosted by the Technical University of Eindhoven in The Netherlands and the University of Genoa in Italy. His interdisciplinary research project combines knowledge of psychology with probabilistic and machine learning techniques to develop a method capable of estimating emotions in crowds based on behavior analysis.

Technical Experience

Probabilistic graphical models, signal processing, machine learning and collective emotion theories.

Current Interests

Affective computing, crowd analysis, human behavior, interactive environments, smart cities, machine learning and deep learning applications.

Funding

EACEA Agency of the European Commission under the EMJD ICE Ph.D. program.

Contact Information

LinkedIn: https://www.linkedin.com/in/oscar-j-urizar/
Personal Email: ossskkar@gmail.com

List of Publications

Journal Articles

1. Oscar J. Urizar, Mirza S. Baig, Emilia I. Barakova, Carlo S. Regazzoni, Lucio Marcenaro, and Matthias Rauterberg. A hierarchical Bayesian model for crowd emotions. Frontiers in computational neuroscience 10 (2016): 63.
2. (Forthcoming) Oscar J. Urizar, Emilia I. Barakova, Carlo S. Regazzoni, Lucio Marcenaro, and Matthias Rauterberg. Emotion Estimation in Crowds: A Machine Learning Approach. (2020).

Conference Proceedings

1. Oscar J. Urizar, Emilia I. Barakova, Carlo S. Regazzoni, and Matthias Rauterberg. Modeling crowds as single-minded entities. In 2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP) (2016), IEEE, pp. 1186-1191.
2. Oscar J. Urizar, Emilia I. Barakova, Lucio Marcenaro, Carlo S. Regazzoni, and Matthias Rauterberg. Emotion estimation in crowds: a survey. In Proceedings of International Conference of Pattern Recognition Systems (2017), IET, pp. 149-154.
3. Oscar J. Urizar, Lucio Marcenaro, Carlo S. Regazzoni, Emilia I. Barakova, and Matthias Rauterberg. Emotion estimation in crowds: the interplay of motivations and expectations in individual emotions. In 2018 26th European Signal Processing Conference (EUSIPCO) (2018), IEEE, pp. 1092-1096.