ESCOLA POLITÉCNICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO
DOUTORADO EM CIÊNCIA DA COMPUTAÇÃO

RODOLFO MIGON FAVARETTO

# **EMOTION, PERSONALITY AND CULTURAL ASPECTS IN CROWDS**: TOWARDS A GEOMETRICAL MIND

Porto Alegre
2019

PÓS-GRADUAÇÃO - *STRICTO SENSU*

Pontifícia Universidade Católica do Rio Grande do Sul

**Pontifical Catholic University of Rio Grande do Sul**
**School of Technology**
**Computer Science Graduate Program**

# EMOTION, PERSONALITY AND CULTURAL ASPECTS IN CROWDS: TOWARDS A GEOMETRICAL MIND

**RODOLFO MIGON FAVARETTO**

Thesis submitted to the Pontifical Catholic University of Rio Grande do Sul in partial fulfillment of the requirements for the degree of Ph. D. in Computer Science.

Advisor: Prof. Soraia Raupp Musse

**Porto Alegre**
**2019**

## Ficha Catalográfica

F272e Favaretto, Rodolfo Migon

Emotion, personality and cultural aspects in crowds : towards a geometrical mind / Rodolfo Migon Favaretto . – 2019.
205 f.
Tese (Doutorado) – Programa de Pós-Graduação em Ciência da Computação, PUCRS.

Orientadora: Profa. Dra. Soraia Raupp Musse.

1. Crowds. 2. Cultural Aspects. 3. Personality and Emotion. 4. Computer Vision. I. Musse, Soraia Raupp. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da PUCRS
com os dados fornecidos pelo(a) autor(a).
Bibliotecária responsável: Salete Maria Sartori CRB-10/1363

Rodolfo Migon Favaretto

# Emotion, Personality and Cultural Aspects in Crowds: towards a Geometrical Mind

This Thesis has been submitted in partial fulfillment of the requirements for the degree of Doctor of Computer Science, of the Graduate Program in Computer Science, School of Technology of the Pontifícia Universidade Católica do Rio Grande do Sul.

Sanctioned on March 29th, 2019.

**COMMITTEE MEMBERS:**

Prof. Dr. Bruno Feijó (DI/PUCRJ)

Prof. Dr. Luiz Chaimowicz (DCC/UFMG)

Prof. Dr. Avelino Francisco Zorzo (PPGCC/PUCRS)

Profa. Dra. Soraia Raupp Musse (PPGCC/PUCRS - Advisor)

# To

everyone who believe and make science happen, especially to those who have contributed with this one.

“If At First the Idea Is Not Absurd, Then There Is No Hope for It.”
(Albert Einstein)

**Acknowledgments**

During these four years, I have only to thank all those who have passed my way and who have certainly left a little of themselves. The moments of joy served to allow me to believe in the beauty of life, and those of suffering served for a unique personal growth. It is very difficult to turn feelings into words, but I will be eternally grateful to you, people who are indispensable for the accomplishment and completion of this thesis.

Firstly, I would like to express my sincere gratitude to my advisor Professor Soraia for the continuous support of my Ph.D. study and related research, for her patience, motivation, and immense knowledge. Her guidance helped me in all the time of research, publishing papers and writing this thesis. I could not have imagined having a better advisor and mentor for my Ph.D at PUCRS.

A very special thank you to Angelo Brandelli Costa and Felipe Vilanova from the Psychology Department, who have always collaborated with my research, always contributing and giving all the necessary support for my research.

I thank my fellow labmates for the stimulating discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had in the last four years. In particular, I am grateful to Leando Dihl, Adriana Braun and Cliceres Dal Bianco for the support, for the fun-time we spent together and for sharing incredible moments at VHLab.

I am also very grateful to my labmates Victor Araújo, Roberto Rosa, Estêvão Testa and Paulo Knob, who helped me during experiments conducted in my research and the sleepless nights that gave us courage to complete tasks before deadlines.

I would like to dedicate a very special thanks to my partner Marcelo, who taught me how to be a better person and "always understood" my moments of absence. A special thanks to my friend Raissa, who always told me words of encouragement and for all the coffee-times we had. Last but not the least, I would like to thank my family: my mom Anadir, my dad Hildo and my brother Cássio for supporting me spiritually throughout the four years of my research and my life in general.

To all of those who have not mentioned, but who have contributed in some way or another to my personal, professional or academic formation, my sincere thanks.

To all of you, my heartfelt thanks!

## Emoção, Personalidade e Aspectos Culturais em Multidões: rumo a uma Mente Geométrica

### RESUMO

O estudo do comportamento humano é um tema de grande interesse científico e, provavelmente, uma fonte inesgotável de pesquisa. A análise de pedestres e grupos em multidões é objeto de interesse em diversas áreas de aplicação, tais como segurança, entretenimento, planejamento de ambientes em espaços públicos e ciências sociais. Aspectos culturais e de personalidade são atributos que influenciam o comportamento pessoal e afetam o grupo ao qual o indivíduo pertence. Neste sentido, a presente tese discute diferentes maneiras de caracterizar indivíduos e grupos em multidões, com o objetivo de propor um modelo computacional para extrair características de pedestres em sequências de vídeos. O modelo proposto considera uma série de características dos pedestres e da multidão, tais como quantidade e tamanho de grupos, distâncias, velocidades, entre outros e realiza o mapeamento destas características em personalidades, emoções e aspectos culturais, considerando as Dimensões Culturais de Hofstede (HCD), o modelo de personalidades Big-Five (OCEAN) e o modelo de emoções OCC. A principal hipótese é que existe relação entre variáveis ditas intrínsecas aos humanos (como emoção) e a maneira pela qual as pessoas se comportam no espaço e no tempo. Como uma das principais contribuições, foram propostas quatro grandes dimensões de características geométricas (Big4GD): *I - Física*, *II - Pessoal e Emocional*, *III - Social* e *IV - Cultural*, as quais buscam descrever o comportamento dos pedestres e grupos na multidão. A ferramenta *GeoMind* foi desenvolvida com o propósito de detectar as quatro dimensões geométricas, a partir de sequências de vídeos. Além disso, diversas análises foram realizadas com o intuito de validar o modelo proposto, desde confrontando resultados com a literatura, incluindo a comparação de multidões espontâneas de diversos países e experimentos controlados envolvendo Diagramas Fundamentais.

**Palavras-Chave:** Multidões, Aspectos Culturais, Personalidade e Emoções, Visão Computacional.

## Emotion, Personality and Cultural Aspects in Crowds: towards a Geometrical Mind


### ABSTRACT

The study of human behavior is a subject of great scientific interest and probably an inexhaustible source of research. The analysis of pedestrians and groups in crowds is an object of interest in several areas of application, such as security, entertainment, planning of environments in public spaces and social sciences. Cultural and personality aspects are attributes that influence personal behavior and affect the group to which the individual belongs. In this sense, the present thesis discusses different ways to characterize individuals and groups in crowds, with the purpose of proposing a computational model to extract pedestrian characteristics in video sequences. The proposed model considers a series of characteristics of the pedestrians and the crowd, such as number and size of groups, distances, speeds, among others, and performs the mapping of these characteristics in personalities, emotions and cultural aspects, considering the Cultural Dimensions of Hofstede (HCD), the Big-Five Personality Model (OCEAN) and the OCC Emotional Model. The main hypothesis is that there is a relationship between so-called intrinsic human variables (such as emotion) and the way people behave in space and time. As one of the main contributions, four large dimensions of geometric characteristics (Big4GD) were proposed: *I - Physical*, *II - Personal and Emotional*, *III - Social* and *IV - Cultural*, which seek to describe the behavior of pedestrians and groups in the crowd. The *GeoMind* tool was developed for the purpose of detecting the four geometric dimensions from video sequences. In addition, several analyzes were carried out with the purpose of validating the proposed model, from comparing results with the literature, including the comparison of spontaneous multitudes from several countries and controlled experiments involving Fundamental Diagrams.



**Keywords:** Crowds, Cultural Aspects, Personality and Emotion, Computer Vision.

# List of Figures

# List of Tables

# List of Acronyms

| | |
|---|---|
| **AE** | United Arab Emirates |
| **AT** | Austria |
| **BR** | Brazil |
| **CN** | China |
| **DE** | Germany |
| **DSF** | Dense Social Forces |
| **DORCA** | Dense Optimal Reciprocal Collision Avoidance |
| **FD** | Fundamental Diagram |
| **FFM** | Five-Factor Model |
| **FPS** | Frames Per Second |
| **ES** | Spain |
| **FR** | France |
| **HCD** | Hofstede Cultural Dimensions |
| **IBM** | International Business Machines |
| **IDV** | Individualism versus Collectivism |
| **IND** | Indulgence versus Restraint |
| **JP** | Japan |
| **LTO** | Long Term Orientation versus Short Term Orientation |
| **MAS** | Masculinity versus Femininity |
| **NMSC** | Non-recursive algorithm of Motion Similarity Clustering |
| **OCC** | Ortony, Clore and Collins |
| **ORCA** | Optimal Reciprocal Collision Avoidance |
| **PDF** | Probability Distribution Function |
| **PDI** | Power Distance Index |
| **PM** | Proposed Method |
| **PT** | Portugal |
| **ROI** | Region Of Interest |
| **SF** | Social Forces |
| **TR** | Turkey |

**UAI** Uncertainty Avoidance Index

**UK** United Kingdom

**VD** Voronoi Diagram

**VO** Velocity Obstacle

**WVMF** Weighted Vector Median Filter

# Contents

# 1. INTRODUCTION

The study of human behavior is a subject of great scientific interest and probably an inexhaustible source of research (Jacques Junior et al., 2010). Due to its importance in many applications, the automatic analysis of human behavior has been a popular research topic in the last decades (Alameda-Pineda et al., 2018). With the improvement of Computer Vision techniques, the detection and tracing of people has become one of the most important areas of video processing. There are many applications such as entertainment (games and movies), understanding of human behavior, security and surveillance, urban planning, activity recognition and planning mass crowd events like sports events or entertainment programs (Shahhoseini and Sarvi, 2019; Zhao et al., 2018).

In particular, the analysis of groups and crowds of people is a phenomenon of great interest in several areas of application (Murino et al., 2017). Without exception, humans in all societies join with other humans to form groups (Forsyth, 2018). When a group of individuals shares the same physical space and has a common goal, they have the characterization of a collective and highly dynamic or, more broadly, a crowd (Li et al., 2012). The formation of a crowd can be observed in various everyday situations, such as people strolling in a mall, the audience that celebrates an event, people attending an airport, among others. Today, with advances in video processing and computing technology systems, it is possible to develop algorithms to detect and identify groups and extract the characteristics of crowds from video sequences (Krüchten and Schadschneider, 2017; Ibrahim et al., 2016; Feng and Bhanu, 2015).

Several work in the literature deal with different applications involving group of pedestrians and crowds, for example, people counting (Kocak and Sevgen, 2017; Kuo et al., 2016; Wang et al., 2015; Chan and Vasconcelos, 2009; Cai et al., 2014), detection of abnormal behaviors (Colque et al., 2015; Yuan et al., 2015; Mahadevan et al., 2010; Zhang et al., 2016), training and movement of groups and crowds (Lu et al., 2017; Solmaz et al., 2012; Zhou et al., 2014; Sethi, 2015; Ma et al., 2017; Jo et al., 2013; Krüchten and Schadschneider, 2017) and detection of social groups in crowds (Solera et al., 2013; Shao et al., 2014; Feng and Bhanu, 2015; Chandran et al., 2015). Most of these approaches are based on people-tracking algorithms or optical flow, and typically consider as characteristics: people's speed, directions, and distances over time.

However, there is one important attribute that can influence the personal behavior that affects the group in which the individual belongs. According to Forsyth (2018), because each society is unique in its traditions and culture, the groups within any given culture may display unique interpersonal processes. Chattaraj et al. (2009) suggest that cultural and population differences can produce changes in the speed, density, and flow of the crowd. In other work, the authors discuss the fundamental diagrams to be used in planning directives (Weidmann, 1993; Predtechenskii and Milinskii, 1978). According to Nazir et al. (2009), culture influences the way people interact with each other and play a large role in the formation of an individual's behaviors and personality, where some aspects can be learned by individuals and others inherited.

Cultural aspects are specific to a group of individuals or category and can be learned by these individuals over time. Interested in these aspects, Geert Hofstede created a model of culturality, which became known as the Cultural Dimensions of Hofstede (HCD) (Hofstede, 2001). It is a model with six cultural dimensions that seek to describe the effects of a society's culture on the values of its members and how these values relate to their behaviors (Hofstede et al., 2010; Hofstede, 2011). Section 2.1.1 gives more details about Hofstede's cultural aspects.

Another factor that can interfere in the cultural aspects of a group of individuals are the personality and emotion of each one. Personality is specific to each individual and can be both learned over time and inherited from ancestors (Hofstede, 2001). When reviewing personality tests, several authors came to the consensus that five factors were sufficient to trace the personality of an individual. These factors became known as the five dimensions of personality *Big-Five* (Costa and McCrae, 1992; Digman, 1990; John, 1990; Goldberg, 1990; Mccrae and John, 1992; McCrae and Costa, 1996). It is a descriptive (taxonomic) psychological model of five factors of personality traits. This model is presented in details in Section 2.1.2. Regarding emotions, Ortony, Clore and Collins's (Ortony et al., 1990) proposed an emotion model called OCC, which is described in Section 2.1.3.

Hofstede and his collaborators (Hofstede, 2011), after a study carried out in more than 30 countries, found that there is a significant link between their model of cultural dimensions and the model of the dimensions of personalities Big-Five (Costa and McCrae, 1992). Thus the culture of a group of individuals and the personality of each are linked. There is a wide variety of individual personalities within the culture of each nation.

The analysis of cultural aspects has not been much explored in the behavioral psychology literature (Berry et al., 2011; Goldstein, 2015), especially when this analysis involves crowds or groups of people. The study of crowds in Computer Science is mainly focused on entertainment and security. However, there are not many papers addressing the cultural aspects of these crowds. Some studies comparing crowds from different countries can be observed in the work of Chattaraj et al. (2009); Lala et al. (2011); Guy et al. (2011); Fridman et al. (2011) and they will be described in Chapter 3.

Following this line of research, this thesis seeks to explore the cultural factors of multi-country crowds from video sequences. The idea is to find characteristics of pedestrians and groups in crowds that are able to describe them and, in a way, differentiate the population from a crowd to another. The goals (general and specific) proposed in this thesis are presented next.

## 1.1 Goals of the Thesis

The main goal of this thesis is to **develop a computational model for detection and analysis of pedestrian and group features in crowds from video sequences**. The idea is to use characteristics of the crowd and their individuals, such as: average distance between people, number and size of groups, speed, density, directions, among other factors that may vary according to the type and nationality of the population to detect and characterize personal, social, emotional and cultural aspects. In particular, we want to propose new factors/dimensions that can be used to characterize crowds, groups and their individuals relating intrinsic features as cultural aspects, personalities and emotions to physical and geometrical manifestations (distances, speeds and so on). Following are listed the specific goals.

### 1.1.1 Specific Goals

In order to meet the main goal, specific goals are proposed, as follows:

- **To create a video database of crowds:** It consists of creating a database of videos of multi-country crowds, containing several types of crowds with different characteristics, such as density, angle of camera and scenarios in which the video was recorded;

- **To define the relevant characteristics for analysis:** In this step, the most relevant characteristics that will be extracted from each video will be defined for further analysis. Some examples are: distances between people, velocities, number and size of groups, densities (people $p/m^2$), among others;
- **To investigate models of culturality, emotion and personality in literature:** This phase was responsible for surveying the literature regarding models that seek to describe the cultural, emotional and personality aspects of groups of people and crowds;
- **To propose factors (dimensions) to describe pedestrian behaviours:** This step is responsible for the definition of new factors to characterize cultural aspects of crowds, also considering the personality and emotions of the individuals. The idea is to define a group of dimensions that are capable of enclose the characteristics of pedestrians in crowds from video sequences relating intrinsic features as cultural aspects, personalities and emotions to physical and geometrical manifestations (distances, speeds and so on); and
- **To develop a software to detect and analyze videos:** Based on the proposed dimensions, a software to automatically detect and analyze pedestrians from video sequences was proposed and developed.

## 1.2 Motivation and Contributions

In the literature, some work seek to simulate different types of crowds, e.g. addressing aspects of gender and age (Kapadia et al., 2013; Pelechano et al., 2016). Computational simulation models that consider, in addition to these factors, data such as culture and habits are still rare (Moore et al., 2008) and require a thorough state-of-art study of other areas of knowledge, such as psychology and anthropology. Reproducing the real characteristics of a crowd in simulation is a challenging task (Allain et al., 2014) and methods that explore cultural aspects in simulated and real crowds are still uncommon.

The motivation for the research in this area starts from the need to have computational tools and models that are capable of extracting characteristics from real pedestrians and crowds, benefiting several other areas of knowledge, e.g. the computational simulation. The relevance of this research is justified due its diverse applications, such as physical space planning, entertainment and security management, among others. These areas seek to consider regional and cultural aspects, but usually this is done with empirical knowledge, there is no computational method or model to help in this task. In the area of physical space management and planning of environments, regional and cultural behaviors and habits should be considered (Hillier, 2002; Nas, 2011). For example, in Arabian cities, the provision of public and private spaces is, for the most part, very different from the European cities, according to an study performed by Hillier (2002).

When it comes to entertainment field, the production of movies and computer games can benefit from the study of cultural aspects. The crowds in the games could be more realistic and include cultural characteristics, typical of each region or country. A crowd of people in a game in Florence in Italy could behave differently from a crowd in a game that takes place in Rio de Janeiro in Brazil, for example. The same goes for the film industry, where pedestrians from simulated crowds could have characteristics and behaviors which are typical of the location of the film's narrative. With respect to the security area, detection of abnormal events could consider cultural aspects, so a strange behavior performed by a pedestrian or a group of pedestrians in France may be different from a strange behavior performed by a pedestrian from Germany.

In order to address this problem, we proposed the development of a computational model that allows the extraction of pedestrian and crowd characteristics based on their behaviours manifestation in space/time, being able to identify cultural, personality and emotional aspects and behaviors that differ from one country to another. Figure 1.1 shows an overview of the features mapping from the proposed approach. Directional arrows indicate which features are used as inputs to calculate other, for example, *distance* is used as input to calculate *socialization* and to detect *groups*.

As one of the main contributions of this thesis, we proposed four different dimensions of features to characterize pedestrians organized in groups/crowds in video sequences. These dimensions, named as **Big-Four Geometrical Dimensions** or just **Big4GD** model, illustrated in Figure 1.1, are:

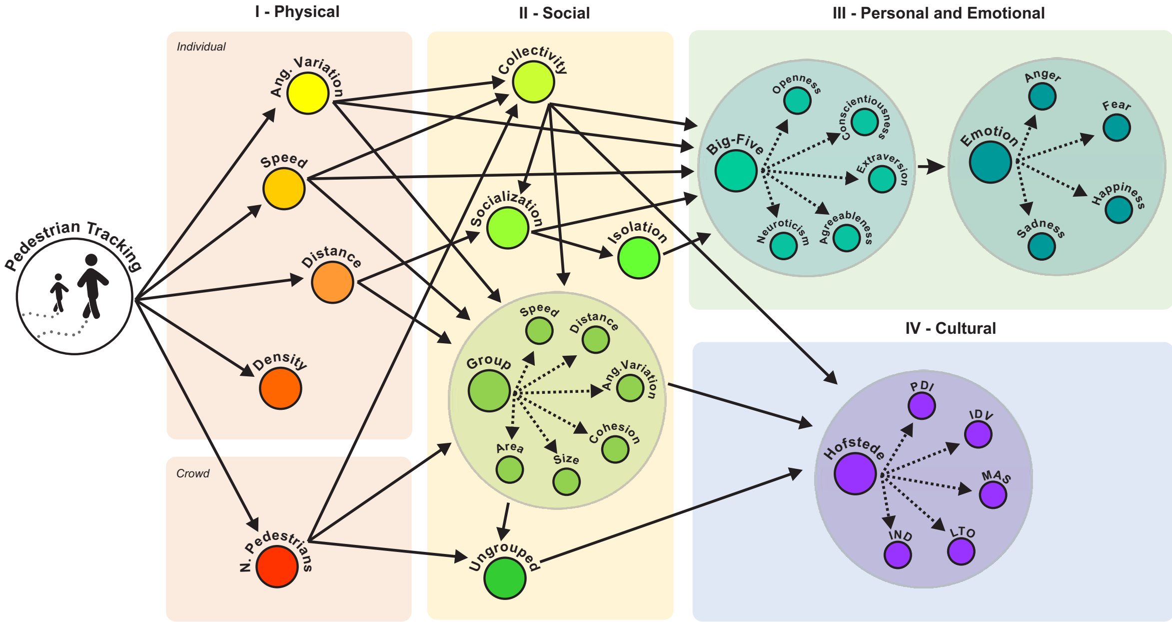



Figure 1.1: **Our Big-Four Geometrical Dimensions**: features mapping from the proposed approach divided into four dimensions: *I–Physical*, *II–Social*, *III–Personal and Emotional* and finally, *IV–Cultural*. Directional arrows indicate which features are used as inputs to calculate others.

- *I – Physical*, which keeps the physical features of pedestrians obtained directly from the tracking, such as speeds, distances from a pedestrian from others, angular variations, among others;
- *II – Social*, which derives from the Physical dimension and deals with social interaction, characterizing groups of pedestrians and social features, as collectivity, isolation and socialization levels of individuals;
- *III – Personal and Emotional*, which maintains the features related to personality (Big-Five) and emotion (OCC) traits;
- *IV – Cultural*, which deals with features regarding cultural aspects, according to Hofstede (HCD).

Naturally, we make no claims regarding the validity of this approach as a scientific tool for assessing cultural, personality or emotional profiles of real people because it is hard to know the intrinsic variables of human beings (personalities/emotions/culture) and to relate that with humans physical/geometrical manifestation. In addition, to get a full validity it should be tested with a a large scale of data, which is very hard to obtain. This approach is a hypothetical model (tested in same examples) to detect our proposed dimensions and its validity in a set of videos as well as some specific literature. Our main hypothesis is that it is plausible to propose a computational model that based on pure geometrical input information can be used to find out some intrinsic information, i.e. correlate geometric manifestations with emotion/personality/cultural aspects. All proposed dimensions and other analysis and experiments, together with the background and related work, are described in next chapters, according to the manuscript structure presented in next section.

A final comment about the contribution of this thesis: we did not find any model in literature that focus on find out and characterizing emotion, personalities and cultural aspects of people in pedestrians from video sequences related to their spacial/time behaviours.

## 1.3 Thesis Structure

The remainder of this thesis is structured as follows: Chapter 2 (*Background*) presents a brief theoretical foundation involving the main areas of knowledge which aims to facilitate the understanding of the thesis. We explore concepts related to crowds and cultural and personality aspects, how these aspects are characterized psychologically, and how they can affect pedestrians in a crowd.

Chapter 3 (*A State-of-the-art Review*) presents the related work identified with main goal to support our research. These work deals with concepts related to crowds, detection of groups of individuals and extraction of information for the mapping of personality, emotion and cultural aspects.

Chapter 4 (*First and Second Dimensions: Data Extraction, Crowd Types and Video Similarity*) shows how pedestrian information and characteristics are extracted from the videos (individual tracking, planar projection, group detection, information extraction) and how that information is used to detect the type of the crowd and similar videos. This chapter basically deals with the features in the *I - Physical* and *II - Social* dimensions, illustrated in Figure 1.1. In addition, this chapter presents some focal results on these specific dimensions.

Chapter 5 (*Third Dimension: Detection of Personality and Emotion Traits*) describes how we performed to detect personality (OCEAN) and emotion (OCC) traits from pedestrians in the video. Chapter 6 (*Fourth Dimension: Detection of *cultural aspects in crowds*) presents how we transformed or mapped pedestrian and group features into Hofstede Cultural Dimensions. These two chapters deals, respectively, with the *III - Personal and Emotional* and *IV - Cultural* dimensions described in Figure 1.1. As before, these chapters also present some obtained results regarding these two dimensions.

Chapter 7 (*Fundamental Diagram Analysis*) shows an analysis of cultural aspects in a controlled experiment, involving Fundamental Diagrams. Considerations about the work developed are made in Chapter 8 (*Final Remarks*). This chapter summarizes the contributions of the research and presents a set of plausible future work.

Finally, Appendix A (*Video Analysis Dataset and Applications*) presents the *Cultural Crowds* dataset, a video database, the software *GeoMind*, developed to detect a series of characteristics from pedestrians and also a viewer which allows to visualize pedestrians features in crowds. Appendix B summarizes the publications obtained during the research of this thesis.

# 2. BACKGROUND

In this chapter we present several concepts and foundations that aim to facilitate and help reading this thesis manuscript, as well as serve as a basis for a better understanding of the next chapters. The content covers crowds and cultural and personality aspects, how these aspects are culturally and psychologically characterized, and how these factors affect pedestrians in a crowd.

Section 2.1 deals with cultural, emotional and personality aspects, presenting the main models in each area. Section 2.2 characterizes crowds and their types and presents the concept of *proxemics*, which deals with interpersonal distances, which may vary across cultures. Section 2.3 presents the Fundamental Diagrams, describing the relationship between speed, density, and flow of pedestrians in a crowd, which may also be influenced by cultural aspects.

## 2.1 Cultural Aspects

Culture, according to Hofstede (2011), is “the collective programming of the mind that distinguishes members of one group of people from another”. This mental programming can take place on three distinct levels: personality, culture and human nature, where some are inherited from ancestors and others are learned by individuals, as illustrated in Figure 2.1.

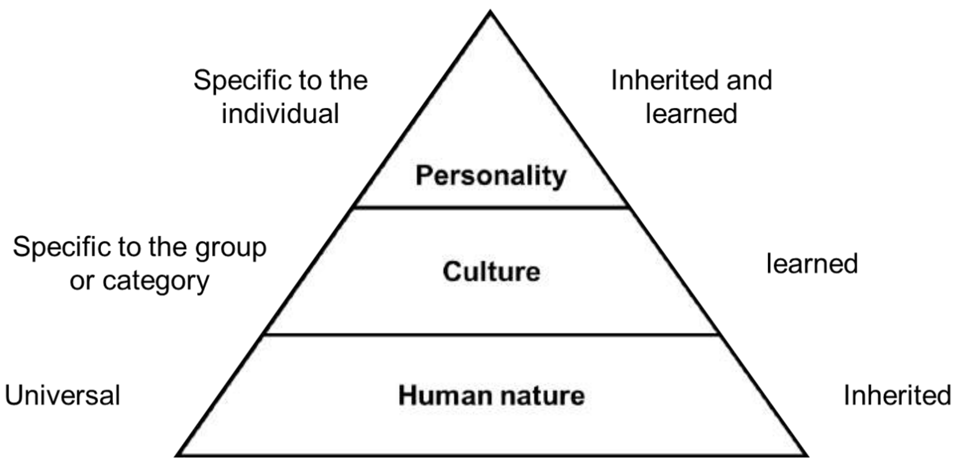



Figure 2.1: **Mental programming levels**: The three levels of uniqueness in mental programming according to Hofstede et al. (2010).

The level of mental programming "Human nature" is universal and, therefore, common to all individuals regardless of geographical location. Considering the purpose of this work, which seeks to investigate cultural aspects in crowds, we studied models that describe, from a psychological point of view, the two levels of mental programming that can be inherited or learned: "Cultural" and "Personality" levels.

At the cultural level, which is specific to a group of individuals or category and can be learned, we studied the model of Cultural Dimensions proposed by Hofstede (2001). At the level of personality, which is specific to each user and can be both learned and inherited, we considered the personality model Big-Five (Costa and McCrae, 1992). According to Revelle and Scherer (2009), personality traits are correlated to emotions of people. Regarding this subject, we studied the OCC (Ortony et al., 1990) emotion model. All this models are described in details in the next sections.

### 2.1.1 Hofstede Cultural Dimensions

Geert Hofstede is an Irish psychologist who, inspired by the anthropologist Clyde Kluckhohn (Kluckhohn, 1962), conducted a large study in the 1970s to investigate the cultural differences between the employees of a multinational company with subsidiaries in several countries (Hofstede, 2001): he analyzed the responses of 117,000 IBM (International Business Machines) employees to understand why their subsidiaries (in Brazil and Japan, for example) were managed differently, despite the efforts of corporate headquarters to implement standardized standards and procedures for all branches.

Hofstede and his collaborators proposed a model of six cultural dimensions, which describe the effects of a society's culture on the values of its members and how these values relate to their behavior, they are: (*i*) Power Distance Index, (*ii*) Individualism versus Collectivism, (*iii*) Uncertainty Avoidance Index, (*iv*) Masculinity versus Femininity, (*v*) Long Term Orientation versus Short Term Orientation and (*vi*) Indulgence versus Restraint Hofstede et al. (1991); Hofstede (2001); Hofstede et al. (2010); Hofstede (2011). Each one of them is described next[1].

[1] Further information on the Cultural Dimensions of Hofstede as well as the indices of various countries can be found at https://www.geert-hofstede.com

### Power Distance Index

Power Distance Index (PDI) is directly related to the organizational hierarchy and represents the degree to which the less powerful members of a society accept and expect power to be distributed unevenly. This index is measured from the point of view of the subordinate, how he deals with inequalities and how he accepts them. A high score at PDI dimension indicates that a society accepts an unequal, hierarchical distribution of power. A low score in this dimension means that power is shared and is widely decentralized and dispersed, also that society members do not accept situations where power is distributed unequally. Table 2.1 shows a summary of the characteristics of PDI dimension at high and low scores.

Table 2.1: **Power Distance Index**: low and high PDI scores characteristics.

| | Characteristics |
|---|---|
| **High** PDI score | Strongly centralized organizations<br>Large wage and status differences<br>More complex hierarchies<br>Large gaps in compensation, authority and respect |
| **Low** PDI score | Flatter organizations<br>More horizontal organizations<br>Supervisors and employees are considered almost as equals |

### Individualism vs. Collectivism

Individualism versus Collectivism (IDV), as the name implies, concerns the degree to which people in society are integrated into groups. Societies with a low degree of individualism tend to privilege the collective, with strong group cohesion, while a high degree of individualism represents societies in which privacy is highly respected, with few interpersonal connections, and very personalized responsibilities. A high IDV score (called Individualism) indicates weak interpersonal connection among those who are not part of a group. People take less responsibility for others' actions and outcomes. A low score in this dimension (called Collectivism) people are supposed to be loyal to the group to which they belong, and, in exchange, the group will defend their interests. Table 2.2 shows a summary of the characteristics of IDV dimension at high and low scores.

Table 2.2: **Individualism vs. Collectivism**: low and high IDV scores characteristics.

| | Characteristics |
|---|---|
| **High** IDV score | Value placed on people's time and need for freedom<br>Promotion is based on rules and competences<br>Expectation of individual rewards for hard work<br>People value the privacy |
| **Low** IDV score | Emphasis on building skills<br>People work for intrinsic rewards<br>Search for harmony overrides moral issues<br>Working for the collective good<br>The promotion considers the context of the individual |

Uncertainty Avoidance Index

Uncertainty Avoidance Index (UAI) expresses the degree to which the members of a society feel uncomfortable with uncertainty and ambiguity. The fundamental question is how a society handles the fact that the future can never be known. Countries with high UAI maintain strict codes of beliefs and behaviors and are intolerant of heterodox behaviors and ideas. Societies with low UAI maintain a quieter attitude, where practice counts more than principles. In societies that score highly for Uncertainty Avoidance, people attempt to make life as predictable and controllable as possible. People in low UAI scoring countries are more relaxed, open or inclusive. Table 2.3 shows a summary of the characteristics of UAI dimension at high and low scores.

Masculinity vs. Femininity

Masculinity versus Femininity (MAS) reflects, to a certain extent, adherence to the separation of values defined by sex: predominantly masculine and feminine functions. The Masculinity side of this dimension represents a preference in society for achievement, heroism, assertiveness, and material rewards for success. Society in general is more competitive. Its opposite, femininity, means a preference for cooperation, modesty, caring for the weak and quality of life. Society in general is more consensual.

In masculine societies, the roles of men and women overlap less, and men are expected to behave assertively, demonstrating your success, and being strong

Table 2.3: **Uncertainty Avoidance Index**: low and high UAI scores characteristics.

| | Characteristics |
|---|---|
| **High** UAI score | Many societal conventions<br>Emotional need for over-effort<br>A high energy society<br>Few risks assumed<br>Very formal conduct in business |
| **Low** UAI score | Extra effort only when needed<br>Informality in business<br>Acceptance of risks<br>Less sense of urgency<br>Openness to changes and innovation |

and fast, are seen as positive characteristics. In feminine societies, however, there is a great deal of overlap between male and female roles, and modesty is perceived as a virtue. Greater importance is placed on good relationships with your direct supervisors, or working with people who cooperate well with one another. Table 2.4 shows a summary of the characteristics of MAS dimension at high and low scores.

Table 2.4: **Masculinity vs. Femininity**: low and high MAS scores characteristics.

| | Characteristics |
|---|---|
| **High** MAS score | Competitiveness and material rewards<br>Distinction between activities of each gender<br>Feelings of pride and importance are attributed to status<br>Money and achievement are important |
| **Low** MAS score | Cooperation and quality of life<br>Power and success equally distributed<br>Relationship oriented/consensual |

Long vs. Short Term Orientation

Long Term Orientation versus Short Term Orientation (LTO) is related to the bonds that a society maintains with its own past in dealing with the challenges of the present and the future. Societies prioritize these two existential goals differently. Low LTO societies, for example, prefer to keep traditions and norms at the same time as they see social change with suspicion. High LTO index cultures, on the other hand, take a more pragmatic approach: they encourage the economy and efforts in modern education as a way of preparing for the future.

Countries with a long-term orientation tend to be pragmatic, modest, and more thrifty. In short-term oriented countries, people tend to place more emphasis on principles, consistency and truth, and are typically religious and nationalistic. Table 2.5 shows a summary of the characteristics of LTO dimension at high and low scores.

Table 2.5: **Long vs. Short Term Orientation**: low and high LTO scores characteristics.

| | Characteristics |
|---|---|
| **High** LTO score | The family is the basis of society<br>Respect for traditions<br>Older people have more power and recognition<br>Virtues and obligations are emphasized |
| **Low** LTO score | Criativity and innovation<br>People open to change<br>Equality in treatment<br>Values and rights are emphasized |

Indulgence vs. Restraint

Indulgence versus Restraint (IND) measures the degree to which each member of society tries to control his/her desires and impulses. Members of more indulgent societies allow themselves to have relatively free and immediate satisfaction of basic and natural human desires, enjoying life and having fun. On the other hand, members of moderate societies are convinced that this bonus must be controlled and regulated by strict rules. Table 2.6 shows a summary of the characteristics of IND dimension at high and low scores.

We did not find any other model which aims to describe cultural aspects regarding different populations. This is why we present only Hofstede Cultural Dimensions model. In the next section we briefly presented the Big-Five personality model.

Table 2.6: **Indulgence vs. Restraint**: low and high IND scores characteristics.

| | Characteristics |
|---|---|
| **High** IND score | Search for happiness and pleasure<br>Less anxiety with deadlines and commitments<br>Importance of freedom of speech<br>Optimism |
| **Low** IND score | Emotions contained<br>Seriousness in meeting deadlines and commitments<br>More controlled and rigid behavior<br>Pessimism |

### 2.1.2 Big-Five Personalty Traits

The model of the five major personality factors, referenced as Big-Five or Five-Factor Model (FFM), is a descriptive (taxonomic) psychological model of five personality traits factors, which were discovered empirically in the early 1980s, accordingly with Goldberg (1982). Researchers Lewis Goldberg, Naomi Takemoto-Chock, Andrew Comrey, and John M. Digman, reviewing the personality tests available at the time, found that most sought to evaluate a subset of five common factors (Goldberg, 1993), which, are known as the five dimensions of Big-Five personality model, also referenced by the acronym OCEAN (Costa and McCrae, 1992; Digman, 1990; John, 1990; Goldberg, 1990; Mccrae and John, 1992; McCrae and Costa, 1996): (*i*) Openness to experience, (*ii*) Conscientiousness, (*iii*) Extraversion, (*iv*) Agreeableness, and (*v*) Neuroticism. Each one of the dimensions is described below:

- **Openness to experience**: represents the degree to which an individual can accept new and unconventional ideas. People who score high on this trait tend to have a wide range of interests. They are curious about the world and other people and eager to learn new things and enjoy new experiences. People with a high score on this feature also tend to be more adventurous and creative. A low score in this trait indicate that people are much more traditional and may have difficulties with abstract thinking;
- **Conscientiousness**: is defined by the self-discipline and obedience which an individual possesses in the performance of a task. People with a high score

on this trait tend to be organized and aware in the details. They plan ahead, think about how their behavior affects others and are on time. A low score indicates that people dislikes structure and schedules and they do not take care of things;

- **Extraversion**: represents how much an individual is being energetic and outgoing, experiencing positive emotions and seeking the company of other people. People who are high in extraversion are outgoing and like to be present in social situations. Being close to other people helps them feel energized and excited. People who have a low score in this trait tend to be more reserved and need to spend energy in social settings. Social events may seem tiresome, and introverts often require a period of solitude and silence to regain their energy;

- **Agreeableness**: determines how compassionate, altruistic and co-operative the individual can be. People who are high in agreeableness tend to be more cooperative while those low in this trait tend to be more competitive and sometimes even manipulative. A high score in this trait indicate that the individuals care about others, feels empathy and concern for other people and enjoy helping and contributing to the happiness of others. Individuals with low score in this trait tend to manipulate others to get what they want, take little interest in others and do not care about how others feel;

- **Neuroticism**: refers to how vulnerable an individual can be by experiencing negative emotional states. Individuals who are high in this trait tend to experience mood swings, anxiety, irritability, and sadness. Gets upset easily, worries about many different things and experiences a lot of stress. Those low in this trait tend to be more stable and emotionally resilient, deal well with stress, rarely feel sad or depressed and do not worry too much.

An instrument widely used to evaluate the personality traits of an individual is the Personality Inventory NEO PI-R (Costa and McCrae, 1992). This test has two versions, one complete and one reduced. In its complete version, it consists of a questionnaire with 240 items that seek to evaluate individuals in the five domains presented through a self-report. Its reduced version (NEO-FFI – NEO-Five Factor Inventory (Costa and McCrae, 1989)) has only 60 items that provide a shorter measure of the five personality domains. In the next section are presented the Big-Five domains and its facets. Each domain has several sub-domains or related categories, which are called facets. For example, one of the 240 items in the NEO PI-R is the statement “I really enjoy talking to other people”, this item is related to the *E1 - Warmth*, which belongs to the Extraversion domain. Each of the five major domains has six related facets, totaling 30. Table 2.7 lists each of the six facets related to each personality domain.

Table 2.7: **Big-Five Personality Traits**: domains and their respective facets – NEO PI-R (Costa and McCrae, 1992).

| Domains | Facets |
|---|---|
| **O**: Openness to experience | O1: Fantasy<br>O2: Aesthetics<br>O3: Feelings<br>O4: Actions<br>O5: Ideas<br>O6: Values |
| **C**: Conscientiousness | C1: Competence<br>C2: Order<br>C3: Dutifulness<br>C4: Achievement Striving<br>C5: Self-Discipline<br>C6: Deliberation |
| **E**: Extraversion | E1: Warmth<br>E2: Gregariousness<br>E3: Assertiveness<br>E4: Activity<br>E5: Excitement-Seeking<br>E6: Positive Emotions |
| **A**: Agreeableness | A1: Trust<br>A2: Straightforwardness<br>A3: Altruism<br>A4: Compliance<br>A5: Modesty<br>A6: Tender-Mindedness |
| **N**: Neuroticism | N1: Anxiety<br>N2: Angry/Hostility<br>N3: Depression<br>N4: Self-Consciousness<br>N5: Impulsiveness<br>N6: Vulnerability |

There are some other models regarding personality traits, such as 16PF (Cattell and Krug, 1986) and Big-Two (Abele and Wojciszke, 2007), but we adopted Big-

Five due to the fact that it is the most known and widely used personality model. In the next section we describe the OCC emotion model.

### 2.1.3 OCC Emotion Model

As mentioned by Revelle and Scherer (2009), “A helpful analogy is to consider that personality is to emotion as climate is to weather. That is, what one expects is personality, what one observes at any particular moment is emotion”. That is the reason why, in our work, we correlated emotion traits with visual behaviors that can be perceived in pedestrian motion.

Ortony, Clore and Collins’s (Ortony et al., 1990) proposed the emotion model called OCC. The OCC emotion model is a widely used model of emotion that states that the strength of a given emotion primarily depends on the events, agents, or objects in the environment of the agent exhibiting the emotion.

The model proposed by Ortony et al. (1990) specifies a set of 22 emotion categories and consists of five processes that define the complete system that individuals follow from the initial categorization of an event to the resulting behavior of the character: a) classifying the event, action or object encountered, b) quantifying the intensity of affected emotions, c) interaction of the newly generated emotion with existing emotions, d) mapping the emotional state to an emotional expression and e) expressing the emotional state. Table 2.8 shows the 22 emotions from OCC model.

There are some other models of emotions in psychological literature, such as Circumplex Model (Russell, 1980) and Lövheims Cube of Emotion (Lövheim, 2012). We use OCC emotion model due to the fact that this model is already used in Computer Vision area. Next section deals with crowds and groups of pedestrians.

## 2.2 Crowds and Groups of People

Considered a very interesting subject, the crowds have been studied by professionals from various fields, such as psychologists, physicists and computer scientists (Kapadia et al., 2013). According to Jacques Junior et al. (2010), the analysis of crowds and groups of people is a phenomenon of great interest in several areas of application (e.g., video surveillance, human behavior understanding

Table 2.8: **Emotion specifications**: emotion type specifications of the OCC model, extract from Ortony et al. (1990).

| Emotion | Specification |
|---|---|
| Joy | (pleased about) a desirable event |
| Distress | (displeased about) an undesirable event |
| Happy-for | (pleased about) an event presumed to be desirable for someone else |
| Pity | (displeased about) an event presumed to be undesirable for someone else |
| Gloating | (pleased about) an event presumed to be undesirable for someone else |
| Resentment | (displeased about) an event presumed to be desirable for someone else |
| Hope | (pleased about) the prospect of a desirable event |
| Fear | (displeased about) the prospect of an undesirable event |
| Satisfaction | (pleased about) the confirmation of the prospect of a desirable event |
| Fears-confirmed | (displeased about) the confirmation of the prospect of an undesirable event |
| Relief | (pleased about) the disconfirmation of the prospect of an undesirable event |
| Disappointment | (displeased about) the disconfirmation of the prospect of a desirable event |
| Pride | (approving of) one's own praiseworthy action |
| Shame | (disapproving of) one's own blameworthy action |
| Admiration | (approving of) someone else's praiseworthy action |
| Reproach | (disapproving of) someone else's blameworthy action |
| Gratification | (approving of) one's own praiseworthy action and (being pleased about) the related desirable event |
| Remorse | (disapproving of) one's own blameworthy action and (being displeased about) the related undesirable event |
| Gratitude | (approving of) someone else's praiseworthy action and (being pleased about) the related desirable event |
| Anger | (disapproving of) someone else's blameworthy action and (being displeased about) the related undesirable event |
| Love | (liking) an appealing object |
| Hate | (disliking) an unappealing object |

or measurements of athletic performance). When a group of individuals shares the same physical space and has a common goal, they have the characterization of a collective and highly dynamic social group or, more broadly, a crowd (Li et al., 2012). A crowd can be characterized as a group of individuals who share the same physical space and can have common goals, or even a collective, highly dynamic social group (Li et al., 2012).

In a crowd, people think they are acting as individuals, but like in other forms of group behaviour, they are being shaped by the collective action of others. People in crowds seem to take on a collective identity, and it may even be difficult to distinguish individual and group behavior. Crowds seem to act as one, even though there may be great diversity within them (Andersen and Taylor, 2007).

A crowd is said to be collective and dynamic (Li et al., 2012), because it can not be treated as a set of individuals with their isolated behaviors, since, in a crowd, individual behavior is influenced by the behavior of his crowd, according to proposed by Pan et al. (2007). From the psychological point of view, the concept of crowd is the concentration of individuals who present behavioral characteristics distinct from those they would present if they were isolated (Bon, 1986).

The formation of a crowd can be observed in various everyday situations, such as people strolling in a mall, the public that celebrates an event, people attending an airport, among others. The study of crowds is not recent, records of research in this area have been found since the early 1970s, as can be seen in Henderson (1971) and Fruin (1971). Regarding computational models we can mention the work proposed by Helbing (1991); Thalmann and Musse (2013); Musse and Thalmann (1997); Still (2000); Helbing et al. (2007); Albeverio et al. (2007); Moussaid et al. (2009); Padgham (2012); Davidich and Koster (2013).

The studies performed by Fruin (1971) shows that the actions taken by a certain crowd are related to the perception of the territory exercised by each of the individuals in this crowd, that is, the way individuals move in the environment and how they position themselves in relation to others is affected by how territorial space is detected and evaluated by each of them. Individuals in a crowd can also be organized as small groups, that is, pedestrians interacting each other (Turner, 1981; Tafjel and Turner, 1986).

According to Forsyth (2010), a group is a social unit whose members stand in status and relationships with one another. Another definition, proposed by Turner (1981), says that a group is composed by two or more people interacting to reach a common goal and perceiving a shared membership, based on both physical (spatial proximity) and social identities.

In this work, groups of individuals are defined based on the distance between people, the speeds in which they move and the direction or goal of each one. In this sense, in order to two or more individuals be part of the same group, they must be near, going to the same direction and moving at similar speeds (or standing). Groups and individuals make up crowds in different configurations, although each individual has its own objectives and behaviour patterns, the behaviour of crowds is widely understood to have collective characteristics, configuring different types of crowds (Davies et al., 1995). In the next section we discuss some types of crowds.

### 2.2.1 Types of Crowds

The behavior of crowds is a widely subject of study and consists of a series of collective characteristics that can be described in general terms. For example, subjective descriptions such as an “angry crowd” or a “peaceful crowd” can be used (Davies et al., 1995). Not distinguish between the crowds can disrupt its control or decision-making, which can result in a series of losses, or even cost lives (in disasters, for example) (Berlonghi, 1995; Kendall, 2016).

A crowd, at a special event, is not just a number of people participating or observing a particular activity (Berlonghi, 1995). A crowd ends up assuming a personality of its own, acting in different ways according to its type. The variability of a crowd's behavior is a consequence of the diversity of the people within it (e.g. age, sex, and social and psychological attributes), as well as the spatial configuration of obstacles and paths (Allain et al., 2014).

These factors have a direct relationship with the characteristics that define the type of a crowd. As proposed by Berlonghi (1995) and Kendall (2016), according to its characteristics, a crowd can be: *i) Casual*, *ii) Conventional* or *iii) Demonstrative*, among other types. Each of these types is presented next.

#### Casual Crowds

*Casual* crowds, according to Kendall (2016), are relatively large gatherings of people who are in the same place at the same time. If there is interaction, it happens only briefly. People in a mall or a bus or subway station are examples of casual crowds. In addition to sharing a momentary interest, such as the presentation of a clown or artist on the street, or even the fall of a small child, a casual crowd has nothing in common.

Another example is the crowd formed by the group of people waiting to cross the street at a busy intersection. In this case, the crowd has a common goal that is only momentary, where the crowd disperses once this goal is achieved, without there being interaction between people. Figure 2.2 shows an example of a casual crowd crossing a street.

Figure 2.2: **Casual crowd**: people crossing the street at Paulista Av. (SP, Brazil), who happen to be in the same place at the same time. Public domain picture extracted from Pexels (https://www.pexels.com).

Conventional Crowds

A *Conventional* crowd is a crowd where a collection of people gather together for a specific purpose (Kendall, 2016). They might be attending a movie, a play, a concert, a party or a lecture. Other examples include religious services and graduation ceremonies. Each of these events have pre-established schedules and standards. Because these events occur regularly, the interaction between participants is much more likely than in *casual* crowds. Crowds leaving social events or environments can also be characterized as conventional crowds.

Goode (1992) thinks that conventional crowds do not really act out collective behavior; as their name implies, their behavior is very conventional and thus relatively structured. Figure 2.3 shows an example of a conventional crowd, a group of people in a party.


Figure 2.3: **Conventional crowd**: a group of people who gather together in a party. Public domain picture extracted from Pexels (https://www.pexels.com).

#### Demonstrative Crowds

*Demonstrative* crowd, according to Berlonghi (1995) is a crowd organized by some leadership, whose actions may include pickets, marches, songs, or demonstrations at specific locations for a specific purpose. Examples are the crowds of people who come together to protest against political, social, cultural, or economic issues. Figure 2.4 shows an example of a demonstrative crowd, a group of people protesting in front a building.

Next section shows an analysis of the space that an individual maintains during interactions with others in a crowd, called *Proxemics*.

### 2.2.2 Interpersonal Spaces

In 1966, the American anthropologist Edward Twitchell Hall (Hall, 1966) proposed the *proxemics* term to describe the use of personal space, area around the individual during interactions and communications with the other individuals of the crowd. An example of *proxemics* is the fact that an individual who finds a seat in


Figure 2.4: **Demonstrative crowd**: a group of people protesting in front of a building. Public domain picture extracted from Pexels (https://www.pexels.com).

a square already occupied by another person at one end tends to sit at the opposite end, preserving a space between the two individuals.

Hall demonstrated that the social distance between individuals may be related to physical distance. In this sense, Hall mentions four distinct zones of interpersonal distances among individuals in a crowd:

- **Intimate distance**: occurs at distances ranging from 15 to 45 centimeters. This level of physical distance often indicates a closer relationship or greater comfort between individuals. It often occurs during intimate contact such as hugging, whispering, or touching. Physical contact is hardly avoided in this area;

- **Personal distance**: occurs at distances ranging from 45 centimeters to 1.2 meters. Physical distance at this level usually occurs between people who are family members or close friends. The closer the people can comfortably stand while interacting can be an indicator of the intimacy of the relationship. At this distance, physical contact occurs with certain ease, being possible to detect facial details and odors;

- **Social distance**: occurs at distances ranging from 1.2 to 3.6 meters. This level of physical distance is often used with individuals who are acquaintances. With someone you know fairly well, such as a co-worker you see several times a week, you might feel more comfortable interacting at a closer distance. In cases where you do not know the other person well, such as a postal delivery driver you only see eventually, a bigger distance may feel more comfortable. Depending on the distance, the physical contact may occur, it is possible to detect some facial details, but it is not possible to detect odors;

- **Public distance**: occurs at distances of at least 3.6 meters. Physical distance at this level is often used in public speaking situations. Talking in front of a class full of students or giving a presentation at work are good examples of such situations. Impersonal and relatively anonymous interactions are perceived, where there is limited sensory involvement. Non-verbal communication is emphasized, including body position, gestures, and movements.

Figure 2.5 illustrates the interpersonal distances proposed by Edward T. Hall. The distance surrounding a person forms a space. The space within intimate distance and personal distance is called personal space (highlighted in orange in the figure). The space within social distance and out of personal distance is called social space (the area highlighted in yellow in the figure). And the space within public distance is called public space (blue area highlighted in the figure).

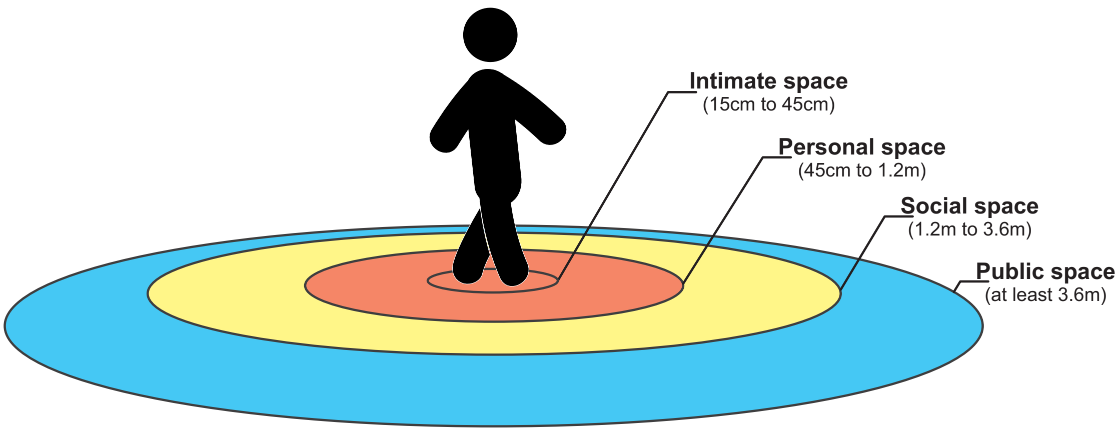



Figure 2.5: **Proxemics**: interpersonal distances proposed by Hall (1966).

Entering somebody's personal space is normally an indication of familiarity and sometimes intimacy (Hall, 1966). However, in modern society, especially in crowded urban communities, it can be difficult to maintain personal space, for example when in a crowded train, elevator or street. Many people find such physical proximity to be psychologically disturbing and uncomfortable, though it is accepted as a fact of modern life. In an impersonal, crowded situation, eye contact tends to be avoided (Hall, 1966).

A person's personal space is carried with him/her everywhere the individual goes. Hall (1966) indicated that different cultures maintain different patterns of personal space. According to him, people from different cultures perceive space (and place) differently. In Latin cultures, for example, relative distances are smaller and people do not feel uncomfortable when they are close to each other. In the Nordic cultures, in contrast, the opposite occurs.

Some attributes can vary from a crowd to another, such as the personal spaces an individual keeps from others, as observed by Hall (1966), and the relation between the speed that individuals move and the density of the crowd (number of people in a certainly area). Next section shows an important instrument used to analysis crowds, the Fundamental Diagram.

## 2.3 Fundamental Diagram in Crowds

Just as cultural aspects can alter people's patterns of interpersonal spaces in a crowd (Hall, 1966), Chattaraj and his co-workers (Chattaraj et al., 2009) suggest that cultural and population differences can also change speed, density, and flow of people in the crowd.

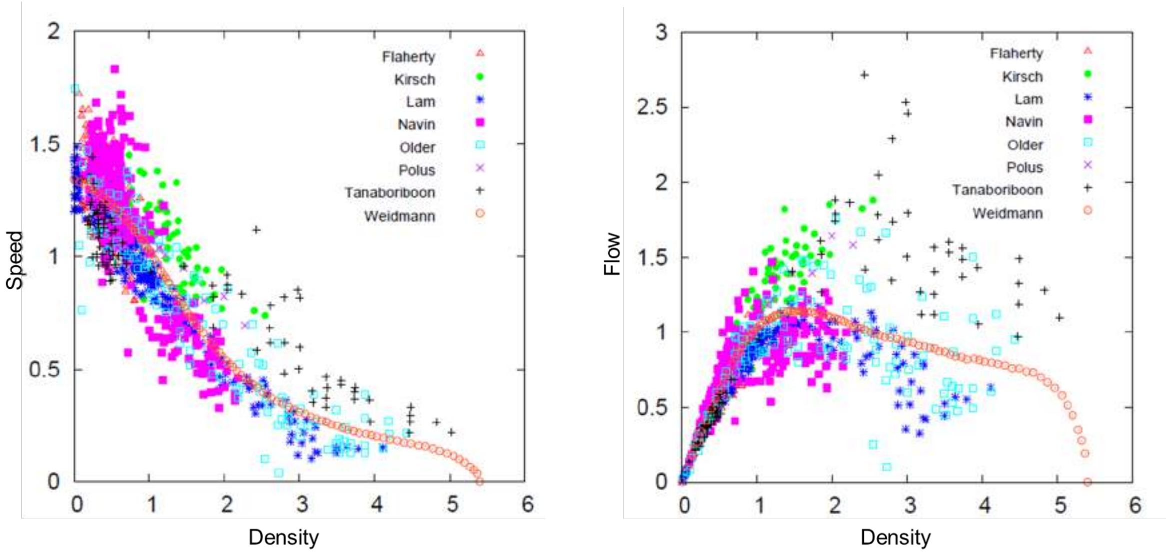
In the work proposed by Chattaraj et al. (2009), the influence of culture on pedestrian trajectory is investigated through fundamental diagrams. The Fundamental Diagrams (FD), originally proposed for use in the traffic planning guidelines (Jelić et al., 2012), are diagrams describing the relationship between three parameters: **density**, **speed** and **flow** (Wu, 2002).

Zhang (2012) also studied Fundamental Diagrams. In his work, FD diagrams have been adapted to describe the relationship between flow and pedestrian density, and are associated with various phenomena of self-organization in crowds such as the formation of pathways (lanes) and jam[2] or overcrowding.

One of the widely used plot is the one who describes the relationship between pedestrian speed and crowd density: as the density increases, the speed decreases (Zhang et al., 2012). Figure 2.6(a), extracted from Zhang et al. (2012), illustrates this idea, comparing several experiments present in the literature involving pedestrian flows, their densities and speeds. Note that the data given by Weidmann

[2]Jam effect occurs when the density of people is so high that the crowd stops moving.

(1993), who performed a meta-study, is an idealized fundamental diagram obtained by collecting and fitting 25 other experiments.




(a) Speed vs. Density (b) Flow vs. Density

Figure 2.6: **Fundamental Diagram**: comparison of fundamental diagrams of bidirectional pedestrian flow from various studies (Zhang et al., 2012).

In terms of the relation between flow and density (Figure 2.6(b)), we can observe the following: as the number of people increases, the density as well as the flow increases. However, as the density increases, it will reach a situation where people will no longer be able to move (jam effect), causing the flow to be zero. There is a density, between zero density and jam density, where the flow will be maximum.

## 2.4 Chapter Remarks

In this chapter we presented a series of concepts involving cultural, personality and emotion aspects, as well as their main classification models that are used in this work. While personality and emotion occur at the individual level, culture is tied to a group or category of individuals. Both personality and culture can be learned by individuals, which entails different cultures and personalities when considering geographically distributed individuals and groups of people, since culture plays a large role in shaping the behaviors and personality of an individual, according to Nazir et al. (2009).

In the sequence, several concepts involving crowds were presented. Among them is the proxemics, which treats the interpersonal spaces around the individuals when they are interacting with others in a crowd. These interpersonal spaces are strongly influenced by the cultural and personality aspects of individuals (Hall, 1966). Just as proxemics suffer cultural influences, the density, velocity, and flow of individuals in a crowd are also affected (Chattaraj et al., 2009). To analyze the relationship between these three factors, the fundamental diagrams (Wu, 2002) are used.

Next chapter presents several work related to the presented concepts, involving crowds, detection of groups of individuals and extraction of information for the mapping of emotion, personality and cultural aspects.

# 3. A STATE-OF-THE-ART REVIEW

In this chapter we present the related work identified with the main goal of this thesis. In crowds, cultural and personality influence can be considered in attributes such as interpersonal spaces, speeds of pedestrians, collision avoidance, formation of groups of people in the crowd, among others. Several researches (Zhan et al., 2008; Weina et al., 2012; Solmaz et al., 2012; Chandran et al., 2015) focus on group identification through computational vision for information extraction. Some group detection and behavior work is presented in Section 3.1.

In addition to information about individuals and their groups, fundamental diagrams can also be used to infer culture. The Fundamental Diagrams describe the relationship between velocity, density and flow of individuals in multitudes, as described in Section 2.3. Section 3.2 presents some work that use FDs in crowds (Cao et al., 2018; Helbing et al., 2007; Seyfried et al., 2010; Wolinski et al., 2014; Narang et al., 2015; Best et al., 2014).

Section 3.3 shows some research about personality (Goldberg, 1982) and emotion traits (Ortony et al., 1990). Finally, Section 3.4 addresses some work that seek to extract or simulate cultural aspects (Hofstede et al., 1991) in crowds (Chattaraj et al., 2009; Guy et al., 2011; Lala et al., 2011; Fridman et al., 2011). Also the work performed by Sorokowska et al. (2017), which investigate the preferred distance that a pedestrian keeps from others in several countries.

## 3.1 Group Detection

An important feature of crowds is their density, and it is natural to think that crowds with different density levels should receive different levels of attention (Zhan et al., 2008). In low and medium-density[1] crowds, some papers address the detection of groups based on trajectories obtained by a tracker or background subtraction techniques. When the crowd presents a high density, an approach for the detection of trajectories based on optical flow is usually applied, since, the identification of groups in this case is not trivial.

[1] In this thesis, it was considered that the density $\langle \Phi \rangle$ (amount of people $p/m^2$) can be: *low*, when $\langle \Phi \rangle \leq 1.5$, *medium*, when $1.5 < \langle \Phi \rangle \leq 4$ and *high*, when $\langle \Phi \rangle > 4$. These values were empirically defined

Weina et al. (2012) proposed an approach to detect groups of individuals who are walking together. The groups are obtained by hierarchical grouping, using the Hausdorff distance (Hausdorff, 1962), considering the proximity and speed of the agents. The distance of Hausdorff, also called Pompeiu-Hausdorff distance, measures how far apart two subsets are from each other. Group identification is based on the bottom-up hierarchical clustering approach, starting with individuals as separate groups and gradually creating larger groups, merging two clusters with the closest proximity to each other (that is, with the shortest distance of Hausdorff).

In the work proposed by Chandran et al. (2015), a non-recursive algorithm of motion similarity clustering (NMSC) is proposed to identify moving pedestrians together in social groups. People are tracked using the background subtraction technique using the Gaussian mixing approach. The clustering algorithm is unsupervised and can automatically identify social groups within a region of interest in a video.

Solera and his collaborators (Solera et al., 2013) propose a new algorithm for the detection of groups, through the clustering of trajectories, solving by means of a parametric group of correlation trained by a support vector machine (SVM). These approachs are mainly aimed at detecting social groups from videos of crowds. Figure 3.1 shows some results obtained in group detection researches in crowds with low and medium densities.

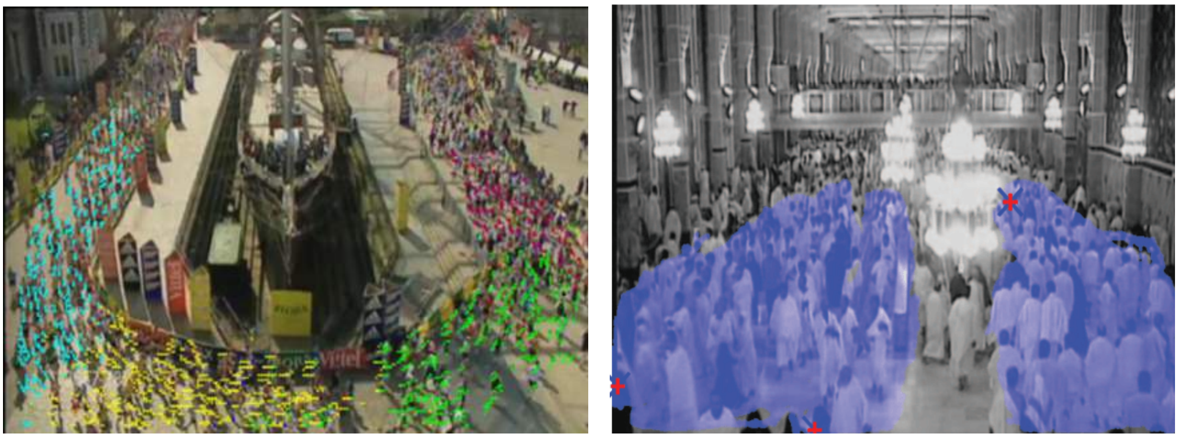
(a) Solera et al. (2013) (b) Weina et al. (2012)

Figure 3.1: **Group detection in low and medium densities**: (a) Solera et al. (2013) and (b) Weina et al. (2012).

Other work, in addition to detecting groups, analyze their behavior (Feng and Bhanu, 2015). Palanisamy and Manikandan (2017) developed a framework for automatic behavior profiling and anomaly detection based on the clustering based group analysis. Some more recent work uses *Deep learning* to analyze crowds (Shao et al., 2015). In some of them, the authors work with detection and recognition of activities in groups of people (Ibrahim et al., 2016; Deng et al., 2016).

However, the detection of groups in high-density crowds is not trivial (Solmaz et al., 2012; Zhou et al., 2014), in which case the focus of the work is the behavior of the crowd as a whole, regardless of group level. Zhou and his collaborators (Zhou et al., 2014) have proposed a collective descriptor for crowds and for the individuals who constitute them. The algorithm *Collective Merging* (Zhou et al., 2014) detects and distinguishes collective motions from random motions.

According to Solmaz and his colleagues (Solmaz et al., 2012), videos of high density crowds make recognition and tracking of individuals and their groups impractical. These authors propose a method where the scene is superimposed by a grid of particles, initializing a dynamic system defined by optical flow. The trajectories of the particles represent the movement in the scene and, through this movement, the behaviors of the crowd are identified. Figure 3.2 illustrates some behavioral detection results in high-density crowds.



(a) Zhou et al. (2014) (b) Solmaz et al. (2012)

Figure 3.2: **Detection of behaviors in high density crowds**: (a) Zhou et al. (2014) and (b) Solmaz et al. (2012).

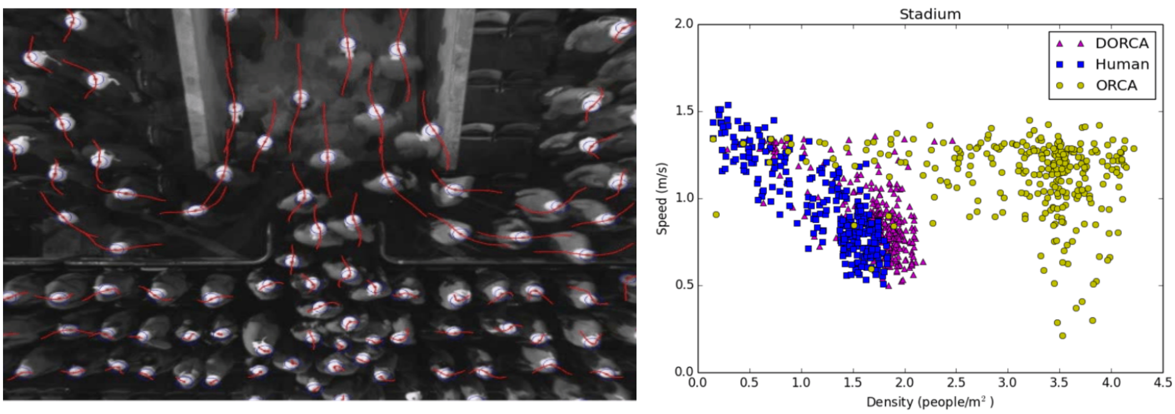
In the next section we present some approaches regarding the use of Fundamental Diagrams in crowds.

## 3.2 Fundamental Diagrams

The understanding of the individual trajectories can arise from studies on the pedestrian dynamics, allowing to analyze the relation between the density of the crowd and the movement of the people. In this sense, the Fundamental Diagrams can serve as an analysis tool (Seyfried and Schadschneider, 2008; Seyfried et al., 2010). FDs can also be used in crowd simulations (Narang et al., 2015; Best et al., 2014).

Helbing et al. (2007) used the FDs to analyze disasters in crowds. In their work, they developed an algorithm to extract pedestrian positions and speeds as a function of time to determine critical conditions in the crowd, which is important for organizing events involving large numbers of people with greater safety.

In the approach proposed by Best et al. (2014), the relationship between flow, speed and density of a real crowd is compared with simulations using the DenseSense algorithm in two versions, each combined with a different local navigation algorithm: (i) *Social Forces* (SF) (Helbing et al., 2000) and (ii) *Optimal Reciprocal Collision Avoidance* (ORCA) (Berg et al., 2009). The two versions of the *DenseSense* algorithm came to be called *DSF* and *DORCA*. Figure 3.3 illustrates one of the experiments carried out in Best et al. (2014).

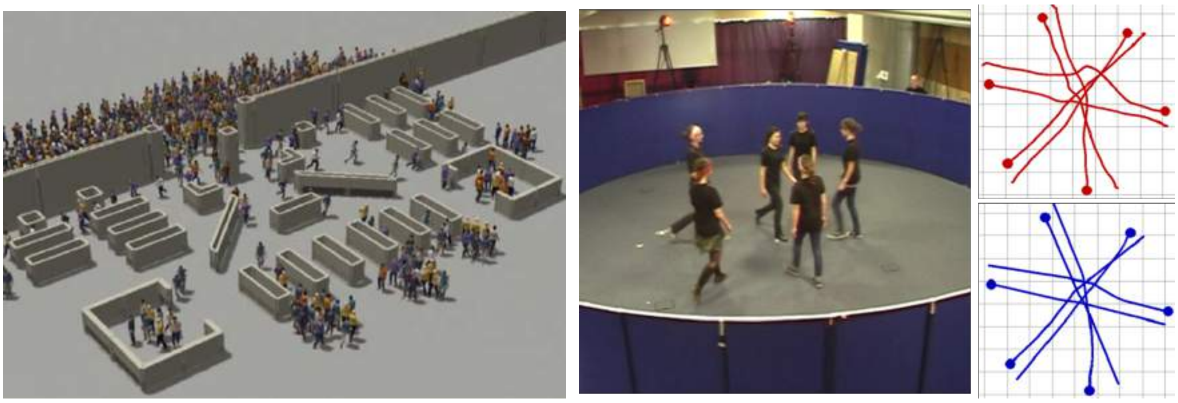



(a) People leaving a stadium (b) FD from the experiment

Figure 3.3: **Experiment of the approach proposed by Best et al. (2014)**: (a) frame of a video of a real world crowd coming out of a football stadium through a tunnel and (b) FD (speed vs. density) from that experiment.

Concerning generation of pedestrian trajectories in crowd simulation, in the work proposed by Narang et al. (2015), movements and behaviors often observed in human crowds were used. The authors proposed the use of FDs to express the relation between speed and density of human individuals in the generation of trajectories in the simulations.

Wolinski et al. (2014) applied FD as a macroscopic data metric for evaluating crowd simulations. The work proposed by Flötteröd and Lämmel (2015) presents a two-way steady pedestrian flow model. Starting from microscopic principles in the simulation, an FD is calculated that defines the specific flow rates of each direction as a function of the specific densities of each direction. Figure 3.4 shows some experiments involving the use of FDs in crowd simulations.



(a) People in a mall (Narang et al., 2015) (b) Trajectory simulation (Wolinski et al., 2014)

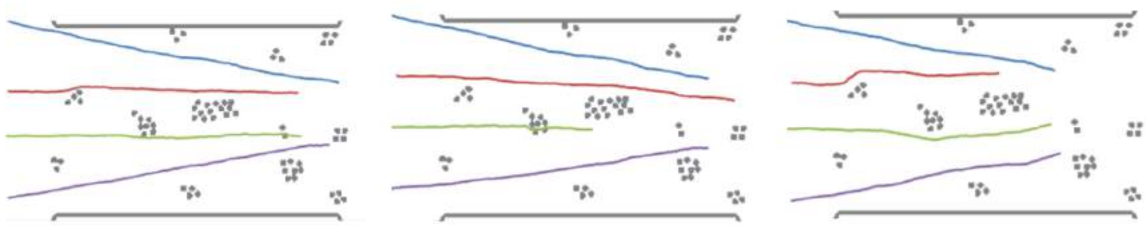
Figure 3.4: **Experiments involving FD and crowd simulations**: (a) Pedestrian simulation in a mall and (b) generation of pedestrian trajectories based on real trajectories (in red the real trajectories and in blue trajectories simulated).

In the next section are presented some work concerning personality and emotion in crowds and pedestrians.

## 3.3 Personality and emotions

The work developed by Guy et al. (2011) presents a new technique for generating heterogeneous behavior in crowds using traits of personalities theory. In that work, the Eysenck 3-Factor personality model was used to establish the range of personality variation. It is a biologically based model based on three independent personality factors: Psychoticism, Extraversion and Neuroticism, also known as PEN, which later inspired the OCEAN model (Big-Five) (Goldberg, 1990).

Figure 3.5 shows some results of the approach proposed by Guy et al. (2011). This figure shows a simulation in a hallway scenario performed by four agents, where the path of each one of them is colored uniquely. This simulation shows a comparison between agents with high levels of (a) “Psychoticism”, (b) “Extraversion” and (c) “Neuroticism”. The high P-factor agents repeatedly cut close to others taking the most direct paths. The high E-factor agents take faster and occasionally “daring” paths, the high N-factor agents take more indirect paths and keep their distance from others.

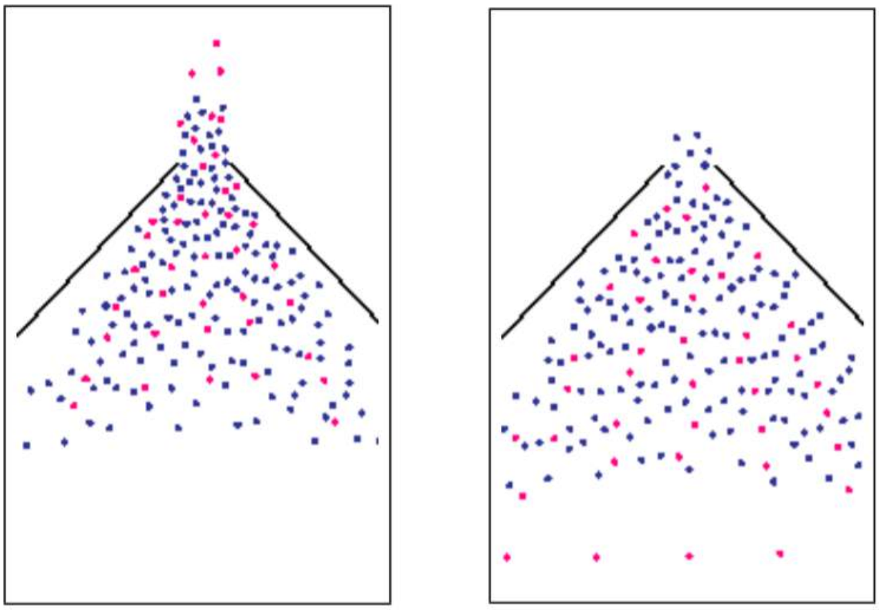


(a) High Psychoticism (P) (b) High Extraversion (E) (c) High Neuroticism (N)

Figure 3.5: **Hallway Scenario simulations performed by Guy et al. (2011)**: a comparison between (a) agents with high levels of *Psychoticism*, (b) *Extraversion* and (c) *Neuroticism*.

Figure 3.6 shows the same time-step from two different simulations. In the left simulation (Figure 3.6(a)), the light red agents are assigned a personality of Aggressive (adjective of Psychoticism). In the right simulation (Figure 3.6(b)), the light red agents are Shy (adjective of Neuroticism). At this point, a few seconds into the simulation, many more Aggressive agents have moved through the exit than the Shy agents. Furthermore, several of the Shy agents can be seen to be holding back away from the exit, causing less congestion.

Saifi et al. (2016) proposed an approach to provide critical emotions in crowd simulators.The simulator uses fuzzy logic for the emotional modeling of critical emotions of members of the crowd at the announcement or the presence of unusual events, in order to quantify emotions. They combine OCEAN (Goldberg, 1990) and OCC (Ortony et al., 1990) emotion model in their approach. Davis and Panksepp (2011) also proposed a similar approach to unifying basic emotions with personality.



(a) Aggressive (b) Shy

Figure 3.6: **Narrowing passage scenario performed by Guy et al. (2011)**: a comparison, at the same time-step, between dark-blue default agents and light-red aggressive agents (a) and light-red Shy agents (b). The aggressive agents exited more quickly, while several Shy agents stay back from the exit causing less congestion.

## 3.4 Cultural Aspects

As explained earlier, the main idea of this work is to find groups in crowds and extract their characteristics, with the purpose of measuring cultural aspects and personalities in videos of crowds from different countries. Some similar work are proposed in the literature, such as those proposed by Lala et al. (2011), Chattaraj et al. (2009) and Guy et al. (2011).

The work of Lala et al. (2011) is quite similar to the one proposed in this thesis, however it uses simulation and not computer vision. The authors introduced a virtual environment that allows the creation of different types of cultural crowds which the users can interact. The parameterization of crowds is based on the cultural dimensions proposed by Hofstede (Hofstede et al., 1991) (presented in Section 2.1.1).

Using one or more cultural dimensions of Hofstede (Lala et al., 2011), the authors proposed an approach to generate crowds with different characteristics and behaviors, proving that the change in crowd type resulted in varied responses by individuals according to the culture. Table 3.1 presents the characteristics and behaviors of the individuals of the simulation attached to each of the dimensions.

Table 3.1: **Hofstede cultural dimensions mapping**: summary mapping of the Hofstede cultural dimensions in behavior and characteristics of the agents in the simulation proposed by Lala et al. (2011).

| Cultural Dimension | Related Features |
|---|---|
| IDV | Interpersonal spaces<br>Walking speeds |
| UAI | Agent navigation<br>Colision avoidance |
| PDI | Personal space given to other agents |
| MAS | Method of treatment of collisions |
| LTO | Steps in collision avoidance algorithm |

As a result of the work proposed by Lala et al. (2011), one of the cultural dimensions, IDV (individualism vs. collectivism) was mapped to agent characteristics and it was found that two distinct types of crowds could be generated. A crowd with a relatively high level of individualism and another crowd representing a more collectivist society.

The agents of the individualistic crowd were characterized with a faster walk and with a greater personal space. Personal space was set at an average of $0.50m$ for the collectivist crowd and $1m$ for the individualistic crowd. The average speed of the individualistic crowd was about 50% faster than the collectivist crowd. These values seem to adequately represent the extreme points of the Hofstede IDV dimension (Hofstede et al., 1991).

Another work which analysis cultural aspects, involves fundamental diagrams. In the work of detecting cultural differences proposed by Chattaraj et al. (2009), a study was carried out to verify the cultural influences in trajectories of individuals, using the FD computed with populations from Germany and India. More specifically, they studied FD in controlled experiments using populations of the same size. Figure 3.7 shows the configuration of the experiment performed.

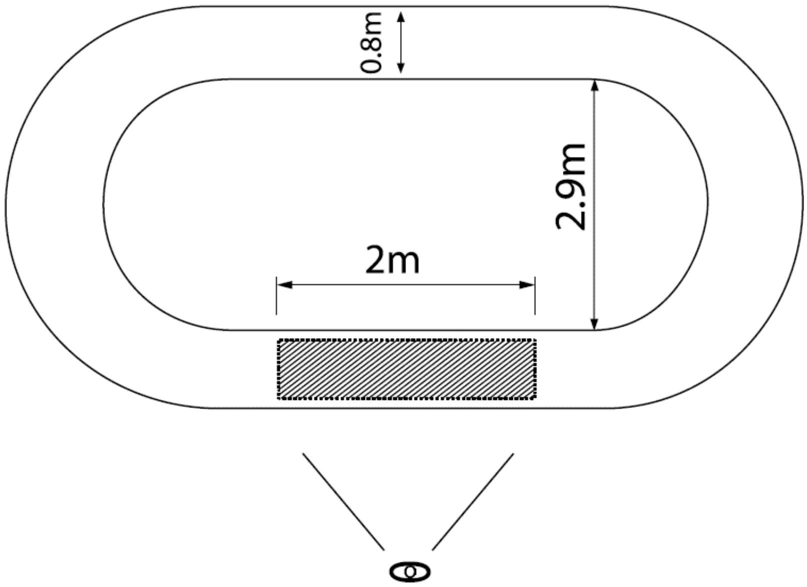
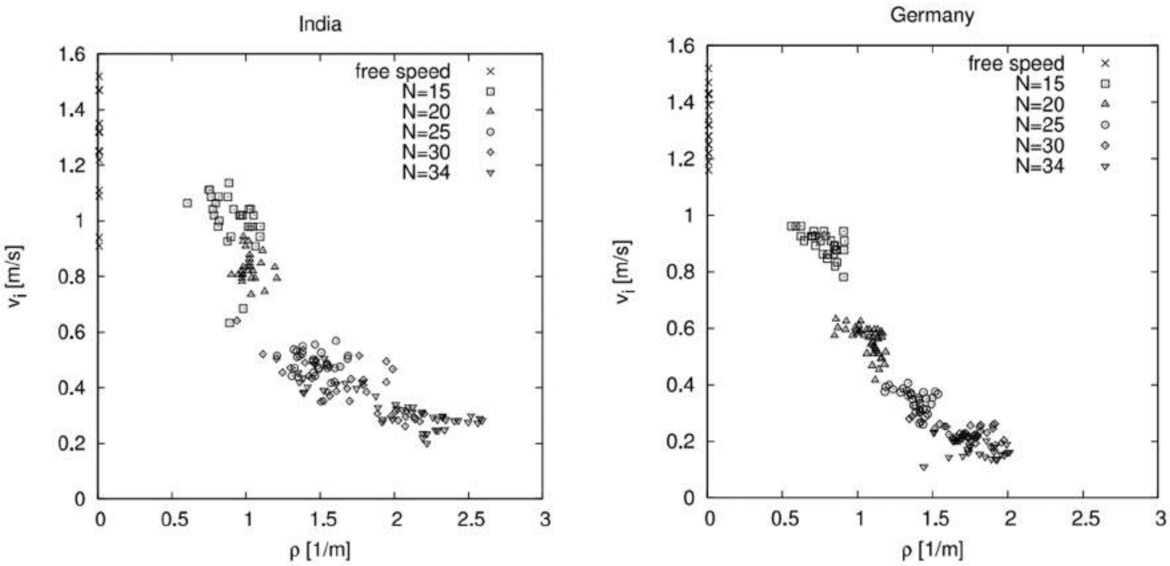



Figure 3.7: **Experiment sketch performed with India and Germany (Chattaraj et al., 2009)**: the length (17.3*m*) and the form of the corridor are identical in both countries.

Figure 3.8 illustrates the FD of each country. The authors observed differences in the minimum personal space estimated for the groups (0.22*m* for India and 0.36*m* for Germany), indicating the influence of the cultural differences of each country. In Germany, speeds are lower than in India, considering the same density of people.

The work proposed by Fridman et al. (2011) presents data that aim to differentiate populations regarding their behavior of movement in crowds. The study is conducted with populations from five countries: France, Iraq, Israel, Canada and England. In this study, cultural parameters are proposed and analyzed in videos from different countries, for later comparison. Some of the parameters analyzed are: speed, personal space, number of collision and population flow. Figure 3.9 shows the average velocities obtained in each analyzed population.

The results presented in Figure 3.9 show that pedestrians in Canada had the highest average speed values. The lowest average speeds were found in Iraq. Based on a visual inspection obtained in the work proposed by Fridman et al. (2011), individuals in Canada move individually, with speeds higher than other cultures. In this same analysis, it was observed that the pedestrians of Iraq are more grouped and with speeds lower than in other cultures.




(a) India (b) Germany

Figure 3.8: **Relationship between speed and density**: speed vs. density of different cultures (Chattaraj et al., 2009): (a) India and (b) Germany. *N* is the amount of participants and free speed is the speed of only one pedestrian walking freely in the corridor. In Germany, speeds are lower than in India, considering the same density of people.

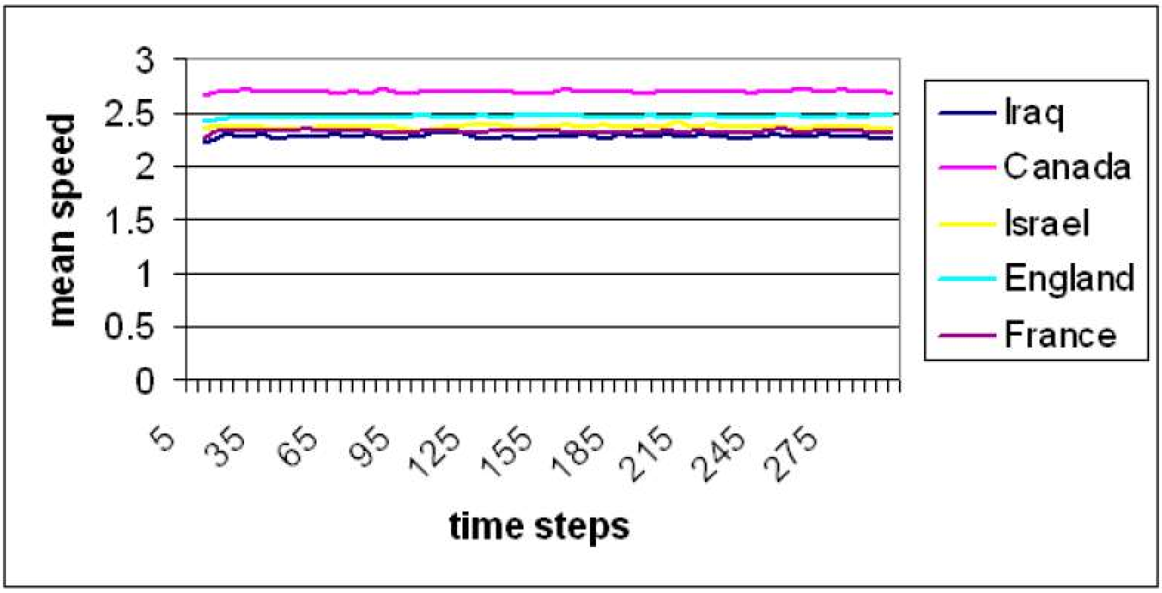



Figure 3.9: **Speed differences**: average speed differences among cultures according to Fridman et al. (2011). Canada had the highest average speed values. The lowest average speeds were found in Iraq.

Another result presented by Fridman et al. (2011) was the difference in pedestrian flow between cultures, as shown in Figure 3.10. The results show that Canada has the highest flow rates, while Iraq, Israel and France have lower flows.

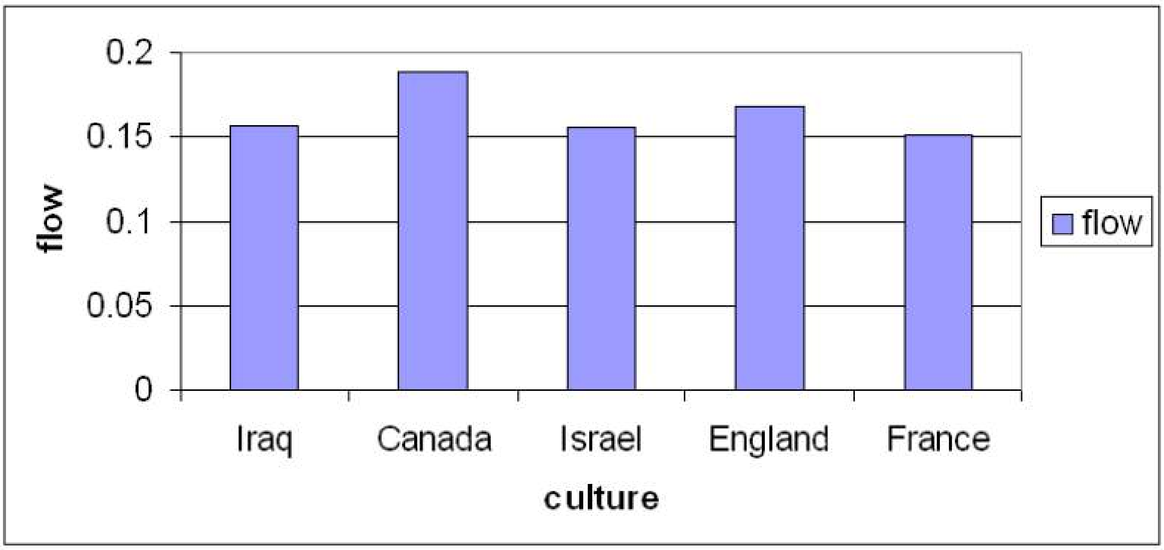



Figure 3.10: **Flow differences**: differences in pedestrian flows between cultures according to Fridman et al. (2011).

Regarding flow of pedestrians, Seyfried et al. (2005) discussed the empirical relationship between the density and velocity of pedestrian movement in a unidirectional flow. Other work (Flötteröd and Lämmel, 2015; Cao et al., 2017, 2018) have addressed bi-directional and multi-directional flows observed at corridors and crossings.

The work proposed by Cao et al. (2018) aimed to investigate differences in pedestrian flows in two cultures: Germany and China. They analyzed the fundamental diagrams of four-directional crossing flow from German and China experiment (See Figure 3.11) and, even though the same measurement method was adopted to analyze the experiments, the difference of the fundamental diagrams from the two countries was very large.

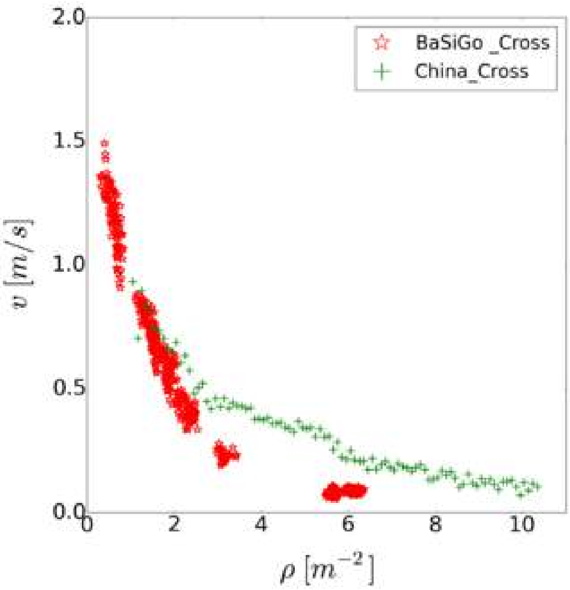
Figure 3.12 shows the FDs from the experiments performed in China and Germany. In the plots presented in Figure 3.12, the red stars indicate the data obtained from Germany (*BaSiGo_Cross*) and the green crosses indicate the data obtained from China (referenced as *China_Cross*). As is showed in Figure 3.12, the specific flow is much larger in Chinese experiment than that in German experiment at the same density, and this is due to higher speed obtained in the crossing area in Chinese experiment.

The authors from Cao et al. (2018) explained this differences saying that in Chinese experiment the participants who are all males are competitive and aggressive, and they move fast to the opposite side when entering into the crossing area.

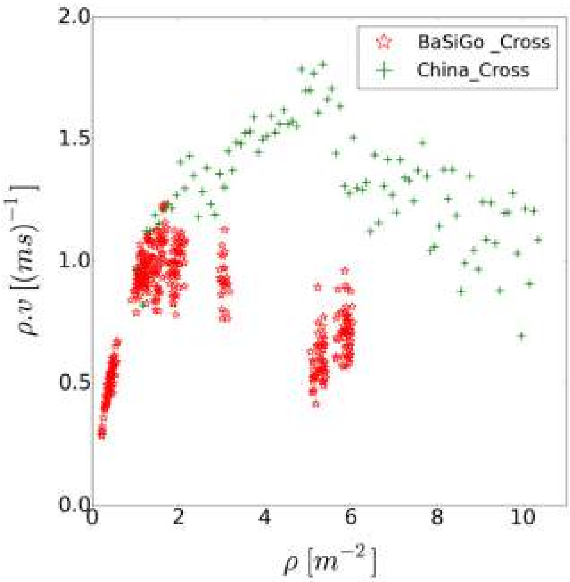
(a) China

(b) Germany

Figure 3.11: **Four-directional cross flow snapshots (Cao et al., 2018)**: snapshots from the experiment performed in (a) China and (b) Germany.




(a) Density–speed




(b) Density–flow

Figure 3.12: **Four-directional cross flow snapshots (Cao et al., 2018)**: snapshots from the experiment performed in China (*China_Cross*) and Germany (*BaSiGo_Cross*).

Even though congestion occurs at high densities, some people still want to move forward and push other pedestrians in the front. However, in Germany experiment pedestrians prefer to keep a comfortable space with other people and no strong pushing or other competitive and varying behaviors are observed at high densities. In other word, the motivation and competitiveness of participants in two experiments are very different, which plays an important role on the large difference between the fundamental diagrams.

Preferred distances that an individual keeps from the others also vary according to the culture (Sorokowska et al., 2017). Interpersonal distance has been categorized (Hall, 1966) in relation to public distance maintained daily between unknown individuals on a street; social distance maintained during formal interactions; personal distance maintained during interactions between friends; and intimate distance maintained in close relationships (Sorokowska et al., 2017).

The study performed by Sorokowska et al. (2017), related to personal distance, was conducted in 42 countries. Participants were asked to complete a visual survey on the amount of distance that they would need to maintain to feel comfortable when interacting with: a) a stranger, b) an acquaintance, and c) a close relation. The authors in turn evaluated projected metric distances for a) social, b) personal and c) intimate distance. In the study, they use a simple graphic task, because it was mostly language independent, anchored by two human-like figures, labeled *A* for the left one and *B* for the right one, as described in Figure 3.13.

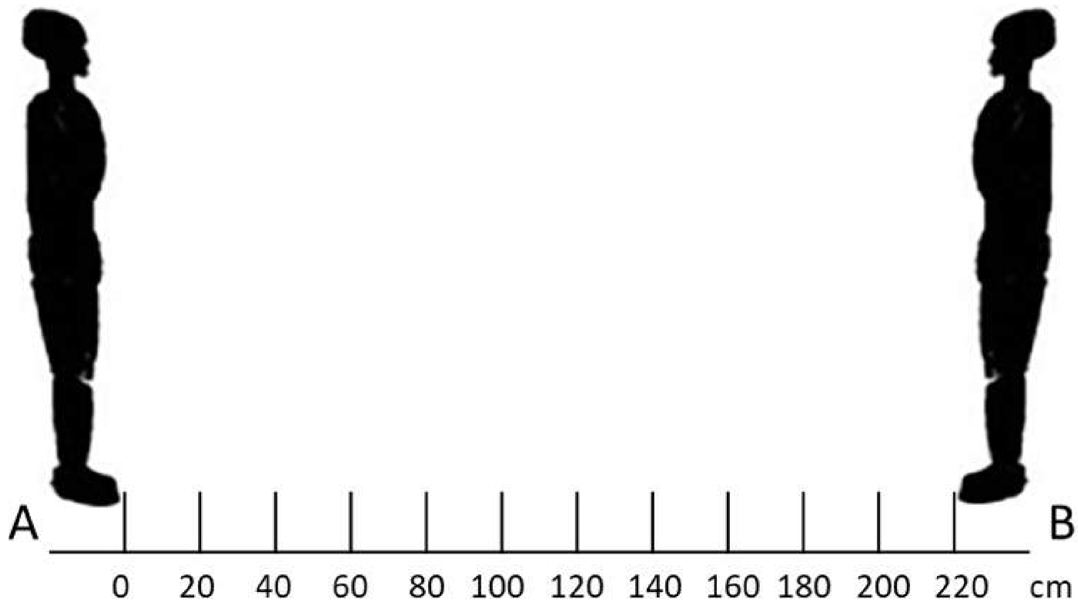



Figure 3.13: **Graphic of distance shown to participants**: distance analysis performed by Sorokowska et al. (2017). Answers were given on a $0 - 220cm$ scale anchored by the two human-like figures. Participants were asked to imagine that he or she is *Person A* and asked to rate how close a *Person B* could approach, so that he or she would feel comfortable in a conversation with *Person B*.

Participants were asked to imagine that he or she is *Person A*. The participant was asked to rate how close a *Person B* could approach, so that he or she would feel comfortable in a conversation with *Person B*. The participants marked the distance at which *Person B* should stop on the scale below the figures. Answers were given on a $0-220cm$ scale anchored by the two human-like figures. Significant variability in personal distance across countries was found from the three different interaction conditions. This work is revisited in the evaluation of our method. We perform an analysis in Chapter 7 (*Fundamental Diagram Analysis*) where we use the approach proposed by Sorokowska et al. (2017). In that analysis, we compared the personal distances measured in our experiments with the distances observed in the Sorokowska's work.

## 3.5 Chapter Remarks

This chapter presented several related work to what is being proposed in this thesis. We sought a literature review that focused on group detection, fundamental diagrams and aspects involving cultural aspects, personalities and emotion traits involving pedestrians or crowds.

Some of the main differentials of the work being presented in this thesis and the related work listed in this chapter are as follows:

- The work presented in this thesis uses, in addition to controlled experiments like those used in the experiments involving the fundamental diagrams, Computer Vision techniques in spontaneous videos of crowds from different countries;
- The work presented in this thesis aims to detect cultural aspects and personality traits of pedestrians based solely on geometric characteristics, such as velocities, distances and angular variation, instead of creating crowds and behaviors in simulated environments using cultural aspects and personalities as inputs.

In the next chapter, we start to present the model proposed in this thesis. Chapter 4 (*First and Second Dimensions: Data Extraction, Crowd Types and Video Similarity*) is responsible for present how the features from the first two dimensions (*I–Physical* and *II–Social*) are defined and computed using physical/geometrical information from the pedestrians tracking.

For that, firstly we present the proposed approach to track the pedestrians in the videos and some initial analysis regarding crowd types and video similarities. It shows how the information and characteristics of the crowds are extracted from the videos (individual tracking, planar projection, group detection, information extraction), how the information is used to carry out the mapping of information extracted to identify types of crowds and videos that look similar to each other.

# 4. FIRST AND SECOND DIMENSIONS: DATA EXTRACTION, CROWD TYPES AND VIDEO SIMILARITY

In this chapter we describe the initial steps of the *Big-Four Geometrical Dimensions* model proposed in this thesis: tracking process, data extraction, crowds classification and video similarity analysis. In this steps, the features from the first two geometrical dimensions (*I–Physical* and *II–Social*) are obtained. Initially, Section 4.1 shows how video information extraction is performed.

The first step in extracting the characteristics of a pedestrian or a crowd in the video is the detection of people and, based on this, the achievement of their trajectories. After that, homogeneous planar projection is performed in order to obtain the parameters in world coordinates (once the trajectories obtained in the tracking process are in image coordinates). Section 4.2 describes the detection of groups and the extraction of their information.

Section 4.3 shows how the similarity analysis of the videos is done. The idea is to find videos with similar characteristics (similar number of people, densities and quantities and sizes of groups) in order to compare and analyze videos containing similar characteristics, in different countries.

In addition, Section 4.4 presents an approach to detect the type of the crowd in the video, according to the classification described in Section 2.2.1 of this manuscript. Finally, Section 4.5 presents the results obtained with this analysis.

## 4.1 Initial Detection and Tracking

The initial detection of people in the video is performed through the real-time object detector proposed by Viola and Jones (2001). This detector, which uses attributes *haar-like*[1] type, had the initial motivation to treat the face detection problem, but it can be used to detect a series of objects.

In this work, to detect the pedestrians, the classifier was trained with 4,500 images of people's heads as positive examples and 1,000 as negative examples. Images for the training were obtained from the images datasets *CoffeBreak* and *Caviar Head* (Tosato et al., 2013). Thus, this detector performs the detection of the

[1] *Haar-like* attributes: also known as *Haar-like features*, are features or fragments of digital images used in object recognition.

initial position of pedestrians in the video based on their heads, as can be seen in Figure 4.1. The head of each pedestrian is detected and marked with a red square, its position is used as an input to the tracking.

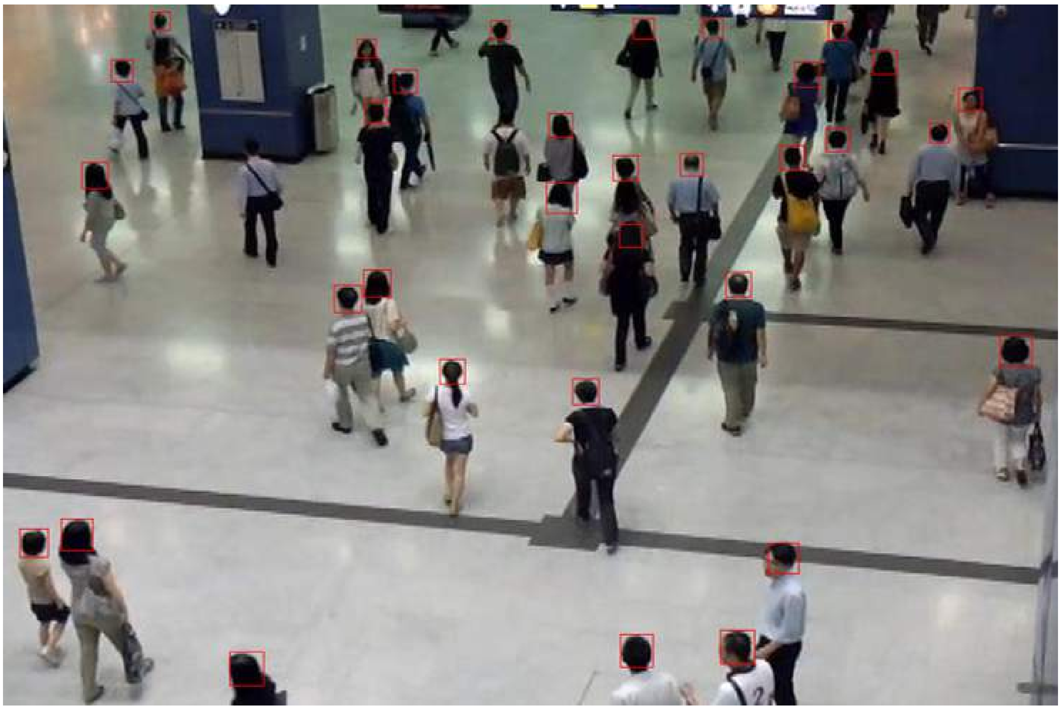


Figure 4.1: **Heads detection**: head detection of pedestrians in the video using a classifier proposed by Viola and Jones (2001).

The Figure 4.2 shows the trajectories (yellow dots indicating the positions of the people in each frame) that each person in the video travels. The trajectories are obtained through the method proposed by Bins et al. (2013). The approach to object tracking is based on several disjoint fragments obtained from the target, which make up a *template*. The fragments are represented parametrically by a mean vector and by a covariance matrix, calculated from a set of characteristic vectors representing each pixel of the target to be traced.

Each of the fragments is independently traced using the distance from Bhattacharyya (Fukunaga, 1990) and the displacement of the *template* is obtained using the Weighted Vector Median Filter (WVMF). To smooth the trajectory and treat short-term occlusions, a displacement vector calculated on the basis of the movement of the target in the previous frames is also used. The appearance changes of the target are handled by a template update scheme.

As mentioned earlier, the tracking input parameters are the initial positions of people's heads. After the tracking step, a set of trajectories, in image coordinates,

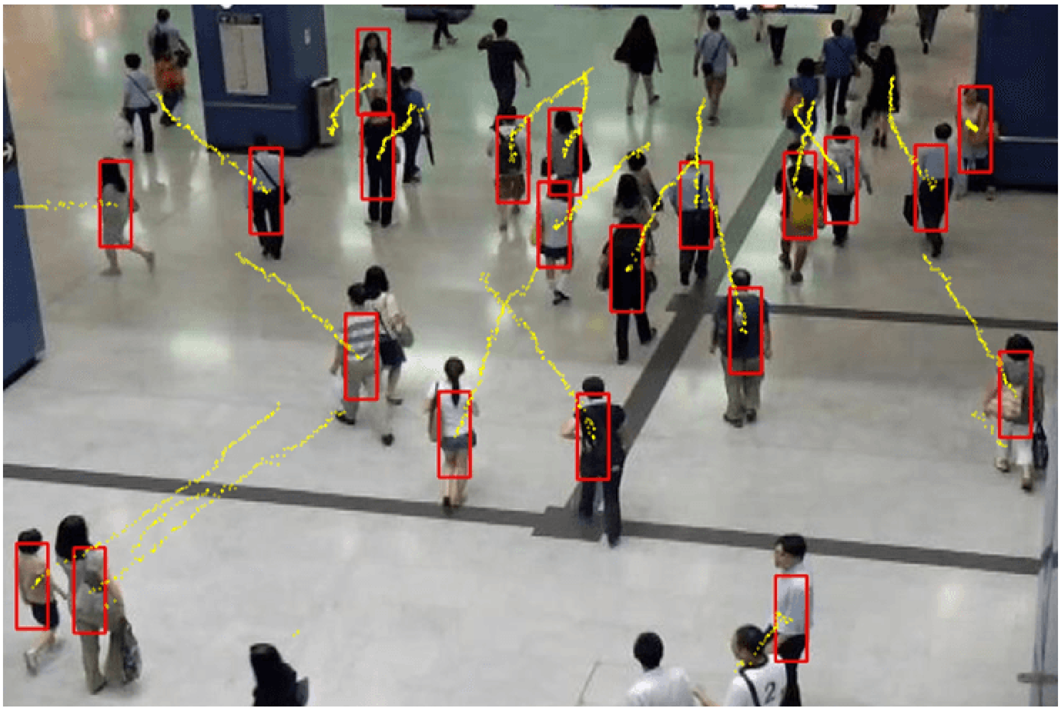


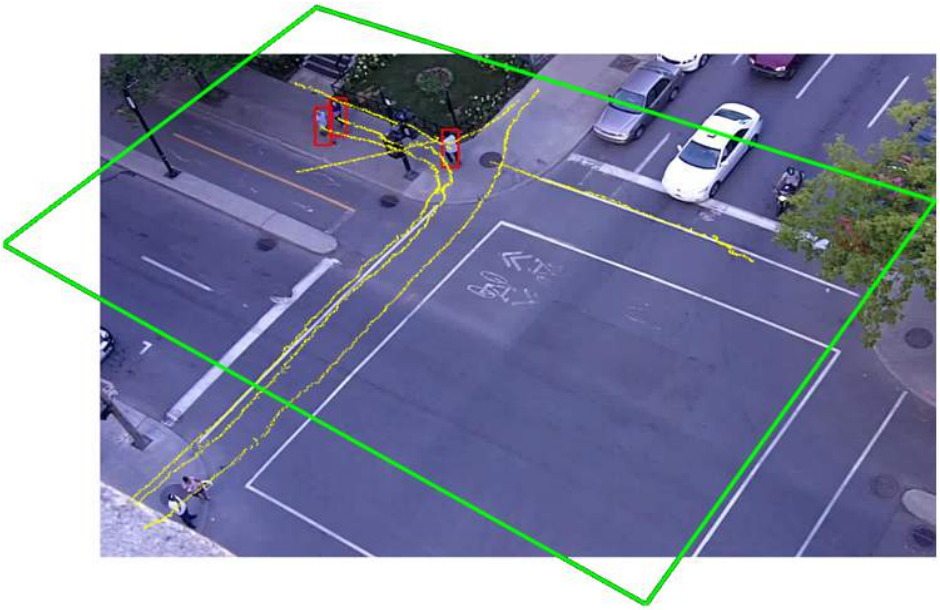
Figure 4.2: **Pedestrian tracking**: detection of people based on their heads and tracking of their trajectories.

is obtained (Figure 4.2). This step can be replaced by any tracker, since the tracking process is not the main contribution of this thesis. The next step is to correct the perspective of the image and obtain the parameters in world coordinates, for this a homographic planar projection is performed, as described in next section.

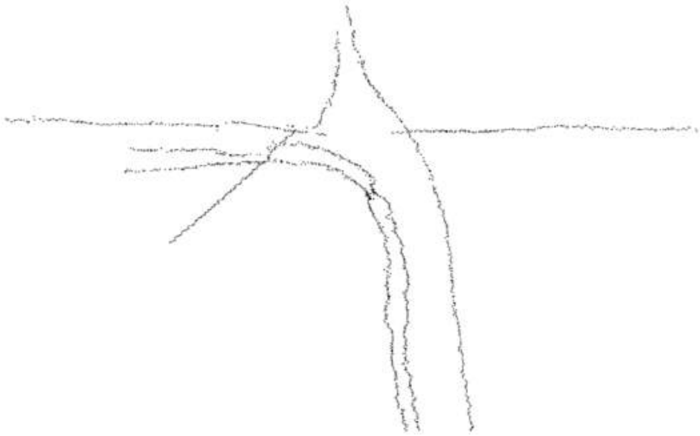
### 4.1.1 Planar Homography

In Computer Vison, planar homography is defined as the mapping of the projection of an object from one plane to another. In this work, the homography is used to correct a perspective image (generated by the angle of the camera that captured the video), generating a planar view of pedestrians' trajectories. Figure 4.3 gives an example of this process: in (a), the trajectories (yellow dots) are in the perspective of the image, which are mapped to an orthogonal plane (b).

In the image coordinate system, it was assumed that the position of people's heads is on the ground (with $z = 0$ in a three-dimensional plane). Since the videos were recorded with the camera positioned in high places and the view of the



(a) Perspective trajectories



(b) Planar projection

Figure 4.3: **planar homography**: trajectories from the perspective of the video camera (a) and after the homography planar projection (b).

people in the crowd is from top to bottom, this assumption does not produce high error in projection.

To obtain the desired parameters in the real world coordinate system, at the time the planar projection is performed, the lateral distance of a person from that video, i.e. the distance between one shoulder and the other, is manually recorded. The average value of this distance in adults is 52*cm*, accordingly to Dreyfuss and Tilley (2002). With this annotation, it is possible to know how many pixels correspond to a meter in that video and thus estimate a series of information, such as the distances between people and the areas occupied by them, necessary in the next steps of the model.

Once the trajectories in world coordinates are obtained, the following information is computed for each pedestrian $i$, tracked at each frame $f$ of the video:

- ($i$) **position** 2D $\vec{X}_i^f = (x_i, y_i)$ (in meters);
- ($ii$) **speed** $s_i^f$ (meters/second): estimated by the distance between the position of the person in the current frame and the position in the previous frame and the elapsed time based on the video's framerate; and
- ($iii$) **angular variation** $\alpha_i^f$ (degrees): obtained through the angle formed by the relation between the position of the person in the current frame and the vector $\vec{r} = (1, 0)$.

Based on these three basic information, the following parameters are calculated for each pair of people $i$ and $j$: ($a$) speed difference $s(s_i, s_j)$, ($b$) orientation difference $\alpha(\alpha_i, \alpha_j)$ ($c$) Euclidean distance $d(\vec{X}_i, \vec{X}_j)$ between the two persons. These parameters are used for the detection of groups of people in the crowd, which is described next.

## 4.2 Group Detection

In this work, groups of individuals are defined based on the distance between pedestrians, the speeds in which they move and the direction or purpose of each one. In this sense, in order to two or more pedestrians to be part of the same group, they must be near, walking in the same direction and at similar speeds (or standing). To define that two pedestrians $i$ and $j$ belong to the same group, they have to attend to three conditions:

- **Condition 1**: if $d(\vec{X}_i, \vec{X}_j) \leq 1.2$ meters. This distance was defined based on the interpersonal space of the proxemics distances proposed by Hall (1966), described in Section 2.2.2 of this volume;
- **Condition 2**: if $\alpha(\alpha_i, \alpha_j) \leq 15°$. It was empirically defined that the difference in orientation between the two pedestrians could not be greater than 15°; and
- **Condition 3**: if $s(s_i, s_j) \leq \beta max\{s_i, s_j\}$. The speed difference between the two pedestrians can not be greater than $\beta$, where $\beta = 5\%$ empirically defined.

Based on these conditions, pedestrians are grouped in pairs and, in a next step, a check is performed to check if two pairs have one pedestrian in common. If they do, they are united into a larger group. This process is performed until the group formations do not share any person, that is, they become disconnected groups.

At first, as soon as a group is detected, it is identified as a group *temporary*, that is, a group that is still unstable. This group becomes a *permanent* group if it maintains its formation (without the input or output of people) for at least 10% of the amount of frames of the video. After this step, a series of information about the permanent groups are noted, in order to be used in the next steps to infer cultural aspects, as listed below:

- Number $G^f$ of groups $g$ at frame $f$;
- Number $n_g$ of members from each group $g$ in the video $v$;
- Number $\nu_v$ of pedestrians who belong to some group in the video $v$;
- Number $\xi_v$ of pedestrians who not belong to any group (pedestrians alone in the video $v$);
- Number $t_g$ of frames in which the group members of $g$ stays together and visible in the video;
- Mean distance $\bar{d}_g = 2\frac{\sum_{i=0}^{n_g-1}\sum_{j=0}^{n_g-1} d(x_i,x_j)}{n_g(n_g-1)}$ (in meters) among all the members of the group $g$;
- Mean speed $\bar{s}_g = 2\frac{\sum_{i=0}^{n_g-1}\sum_{j=0}^{n_g-1} s(s_i,s_j)}{n_g(n_g-1)}$ (em meters/second) among all the members of $g$;
- Mean angular variation $\bar{\alpha}_g = 2\frac{\sum_{i=0}^{n_g-1}\sum_{j=0}^{n_g-1} o(\alpha_i,\alpha_j)}{n_g(n_g-1)}$ (in degrees) among all the members of $g$;
- Group area $A_g$. This calculation is detailed in the next section; and
- Group cohesion $C_g$, which is detailed in the Section 4.2.2.

### 4.2.1 Group Area

The area of the group $A_g$ is calculated based on the number of members: if the group has at least three members, the area of the polygon formed by the pedestrians in the ground is calculated, where each of its vertices is formed by the position of each member. If the group has only two members, the area is calculated based on the rectangle formed between them. The width of this rectangle is the average circumference of an adult person (80 centimeters, accordingly to Dreyfuss and Tilley (2002)) and the length is the distance between the members. Figure 4.4 illustrates this idea.

(a) Two members (b) Three or more members

Figure 4.4: **Different cases to obtain the group area**: group with only two members (a) and a group with three or more members (b).

### 4.2.2 Group Cohesion

The $C_g$ cohesion arises with the strong connection between members of a social group. Members of strongly cohesive groups are more inclined to remain in the group (Dyaram and Kamalanabhan, 2005). Our method for group cohesion is inspired by the work of Bassi (2006), who propose that how much a person wants to remain in their group can be quantified through the stability of the relationships within that group. The cohesion $C_g$ of a group is calculated as described in Equation 4.1:

$$C_g = \frac{1}{n_g} \sum_{i=1}^{n} C_i, \tag{4.1}$$

where $C_i$ represents the cohesion of the individual $i$ in a given group $g$ and $n_g$ is the number of members of the group $G$. The cohesion $C_i$ of a determined pedestrian is given by $C_i = CO_i + CV_i + CP_i$, where $CO_i$ is the cohesion related to the orientation of the individual, $CV_i$ is the cohesion related to the speed of this individual and $CP_i$ is the population cohesion, related to the size of the group. These three factors of cohesion are calculated through the following empirically defined equations:

$$CO_i = \frac{w - |\bar{\alpha}_g - \alpha_i|}{w}, \tag{4.2}$$

$$CV_i = \frac{\gamma - |\bar{s}_g - v_i|}{\gamma}, \tag{4.3}$$

$$CP_i = \log n_g. \tag{4.4}$$

In the orientation cohesion $CO_i$ (Equation 4.2), the constant $w = 180^\circ$ indicates the angle for minimum orientation cohesion. In the velocity cohesion $CV_i$ (Equation 4.3), the constant $\gamma = 1.4m/s$ is equivalent to the maximum speed variation related to the minimum velocity cohesion. In population cohesion $CP_i$ (Equation 4.4), $n_g$ corresponds to the amount of group members of $g$. Cohesion $C_g$ (Equation 4.1) is calculated for each frame for each group present in the video. This information, along with the others presented in this chapter, are used in the next steps to determine similarity in videos and infer cultural aspects.

Group detection analysis

Figure 4.5 shows a frame of a video from China with some groups detected using our approach: the group highlighted in blue is the temporary group. This group was detected some frames ago when that three pedestrians entered in the video. If that pedestrians remain in the formation for at least 10% of the frames of the video, they will become a permanent group (as the other five groups highlighted in red in Figure 4.5).

In the next section is presented a video similarity analysis, an approach to detect videos with the same characteristics, in order to make a proper comparison between them.

Figure 4.5: **Group detection example**: six groups detected, being five permanent, in red, and one temporary, in blue.

## 4.3 Similarity Analysis

We propose a way to find similar videos. With the goal of making a proper comparison between different videos (for example, it would not be appropriate to compare a video with some carefree people strolling in the mall with a video where there are many people at a train station, hasty, and preoccupied with the time, since the context is different). For each video $v$, is calculated, from all the frames of $v$, the following information: (*i*) number of groups $G_v$, (*ii*) number of pedestrians who belong to any group $\nu_v$ and (*iii*) number of pedestrians who are ungrouped $\xi_v$. Also, the percentage $\tau_v$ of frames in the video $v$ which contains at least one group is noted.

For the detection of similarity in videos, an extension of the Fundamental Diagrams, presented in Section 2.3, is used. In this work, instead of the data commonly used in FD (density $\times$ flow), information about crowds (density, velocity and collectivity) obtained in each measurement area (*ms*) is used, which is a part of the image empirically defined as a region of $6m^2$ ($\Delta x = 3m$ e $\Delta y = 2m$), as illustrated by Figure 4.6. As the crowds used in the experiments are of low and medium densities, we have chosen to use areas *ms* greater than $1m^2$, thus avoiding that many regions have zero density.

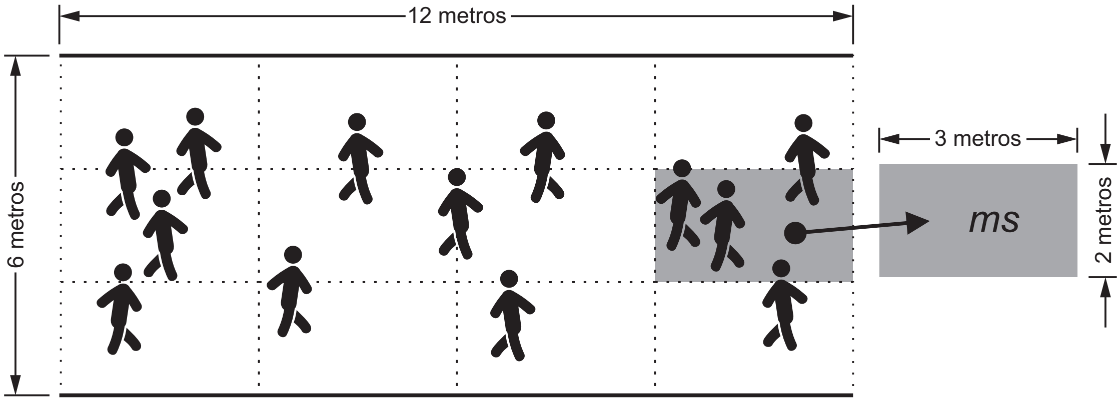



Figure 4.6: **Measurement areas**: sub-regions *ms* of the image where density, velocity, and collective information are obtained.

Consequently, in each region *ms* of the image, three types of information are calculated: density $\langle\Phi\rangle$, speed $\langle\Theta\rangle$ and collectivity $\langle\Psi\rangle$. The density $\langle\Phi\rangle$ (people/$m^2$) is obtained by dividing the number of people by the *ms* area gives at each frame, as defined by Zhang et al. (2011):

$$\langle\Phi\rangle_{ms} = \frac{N_a}{\Delta x.\Delta y}, \tag{4.5}$$

where $N_a$ is the number of people in *ms* and $\Delta x$ and $\Delta y$ are, respectively, the length and width of *ms*. The average density $Avg\langle\Phi\rangle^f$, at each frame $f$, is then calculated as follows:

$$Avg\langle\Phi\rangle^f = \frac{1}{N_{ms}}\sum_{o=1}^{N_{ms}}\langle\Phi\rangle^f_{ms_o}, \tag{4.6}$$

where $N_{ms}$ is the number of measurement areas *ms* in which the image was divided.

The speed $\langle\Theta\rangle$ ($m/s$), also inspired by Zhang et al. (2011), is the average of the instantaneous velocities $s_i$ from all the pedestrians $i$ in the measurement area *ms*, given by:

$$\langle\Theta\rangle_{ms} = \frac{1}{N_a}\sum_{i=1}^{N_a} s_i, \tag{4.7}$$

where $N_a$ is the number of pedestrians in the measurement area *ms* and $s_i$ is the instantaneous velocity of pedestrian $i$, computed as follow: $s_i = \frac{\delta d}{\delta t}$, where $\delta d =$

$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$, $(x_1, y_1)$ e $(x_2, y_2)$ are the 2D initial and final positions of a pedestrian at period $\delta t = 1s$. At each frame $f$, the average speed $Avg\langle\Theta\rangle^f$ is defined by the Equation 4.8:

$$Avg\langle\Theta\rangle^f = \frac{1}{N_{ms}} \sum_{o=1}^{N_{ms}} \langle\Theta\rangle^f_{ms_o}, \tag{4.8}$$

where $N_{ms}$ is the number of measurement areas $ms$ at frame $f$.

The collectivity $\langle\Psi\rangle$, computed every two pedestrians $i$ and $j$, was inspired by Zhou et al. (2014). However, as in this work the goal is to measure the collectivity at each frame, without considering the future path of the pedestrians, it is not considered the similarity factor proposed in the work of Zhou et al. (2014). Thus, the collectivity between two pedestrians $i$ and $j$ is calculated as a decay function as $\varpi(i,j) = s(s_i, s_j).w_1 + \alpha(\alpha_i, \alpha_j).w_2$, considering $s$ and $\alpha$, respectively, the speed and orientation differences between the two pedestrians $i$ and $j$, and the parameters $w_1$ (in meters) and $w_2$ (in radians) are constants that must regulate the equation. The values defined in this work are $w_1 = 1$ and $w_2 = 1$. Thus, the values of $\varpi(i,j)$ belong to the range $0 \leq \varpi(i,j) \leq 4,34$. The collectivity in a measuring area $ms$ is defined by:

$$\langle\Psi\rangle_{ms} = \frac{1}{N_a^2} \sum_{i=1}^{N_a} \sum_{j=1}^{N_a} \gamma e^{(-\beta\varpi(i,j)^2)}, \tag{4.9}$$

where, again, $N_a$ is the number of pedestrians in that $ms$, $\gamma = 1$ is the maximum value of collectivity when $\varpi(i,j) = 0$, and $\beta = 0.3$ is empirically defined as the decay value. Consequently, $\langle\Psi\rangle_{ms}$ is a value in the interval $[0; 1]$. The average collectivity $Avg\langle\Psi\rangle$, computed at each frame $f$ is given by:

$$Avg\langle\Psi\rangle^f = \frac{1}{N_{ms}} \sum_{o=1}^{N_{ms}} \langle\Psi\rangle^f_{ms_o}. \tag{4.10}$$

At this point, for each video $v$ it is obtained a vector $\vec{V_v}$ of computed data, where $\vec{V_v} = [G_v, \nu_v, \xi_v, \tau_v, Avg\langle\Phi\rangle_v, Avg\langle\Theta\rangle_v, Avg\langle\Psi\rangle_v]$ represent, respectively, number of groups, number of grouped pedestrians, number of ungrouped pedestrians, percentage of frames in the video which contains groups, average density of the video, average speed and average collectivity of pedestrians. Each element of $\vec{V_v}$ is quantified in one of three values: low, medium and high, in order to provide ratings for the crowds in the videos. The quantification is performed through the clustering process.

In order to quantify each element in $\overrightarrow{V_v}$, the clustering algorithm *K-means*, proposed by MacQueen (1967), is used. More precisely, the value of $k = 3$ is defined to generate exactly three clusters, and the *K-means* algorithm is applied to each individual element of $\cup_v \vec{V}_v$, this is the set of vectors corresponding to all videos analyzed. The centroids of each class of the *K-means* algorithm are then used to classify each element of $\overrightarrow{V_v}$ into low, medium or high, respectively called $S_0$, $S_1$, and $S_2$. For example, a particular video $v$ may have a small number of groups ($G_v \in S_0$), a high number of individuals grouped ($\nu_v \in S_2$) and an medium value of individuals which do not belong to any group ($\xi_v \in S_1$). In addition to the information used in the quantization process, the standard deviation for all average values ($Std\langle\Phi\rangle, Std\langle\Theta\rangle, Std\langle\Psi\rangle$) is also calculated.

In order to carry out the characterizations of the crowds and the analysis of similarity, besides the information extracted from the videos, the types of crowds presented in Section 2.2.1 of this manuscript were considered. Thus, the information considered in this analysis is:

- Crowd type (Casual, Conventional and Demonstrative);
- Presence of groups ($G > 0$);
- Size of groups (no groups, medium and large), based on $\nu_g$;
- Crowd density (low, medium and high), based on $Avg\langle\Phi\rangle_v$; and
- Interaction level in the crowd (low, medium and high), based on $Avg\langle\Psi\rangle_v$.

The listed parameters are mapped directly to specific elements of $\overrightarrow{V_v}$, except for the type of crowd. In the next section is presented the proposed approach to automatically detect the crowd type (Casual, Conventional and Demonstrative).

### 4.4 Crowd Types Classification

To obtain the type of crowd, three hypotheses, one for each type (casual, conventional and demonstrative), empirically defined were proposed. Each one of them is presented below:

- **Casual** (**H1**): (*i*) low or medium density of pedestrians $Avg\langle\Phi\rangle_v \in S_0$ or $S_1$, (*ii*) pedestrians walking at low or medium speeds $Avg\langle\Theta\rangle_v \in S_0$ or $S_1$, (*iii*) small or medium amount of groups $G_v \in S_0$ or $S_1$, (*iv*) low collectivity $Avg\langle\Psi\rangle_v \in S_0$ and (*v*) low frequency of groups $\tau_v \in S_0$;
- **Conventional** (**H2**): (*i*) medium or high density of pedestrians $Avg\langle\Phi\rangle_v \in S_1$ or $S_2$, (*ii*) pedestrians walking with high speed and low variation $Std\langle\Theta\rangle_v \in S_0$, (*iii*) small groups $G_v \in S_0$, (*iv*) most of pedestrians walking alone $\xi_v \in S_2$ and (*v*) medium or high collectivity $Avg\langle\Psi\rangle_v \in S_1$ or $S_2$;
- **Demonstrative** (**H3**): (*i*) high density of pedestrians $Avg\langle\Phi\rangle_v \in S_2$, (*ii*) pedestrians walking at low speed $Avg\langle\Theta\rangle_v \in S_0$, (*iii*) high collectivity $Avg\langle\Psi\rangle_v \in S_2$, (*iv*) most of pedestrians walking alone $\xi_v \in S_2$ and (*v*) high frequency of groups $\tau_v \in S_2$.

High-level crowd characterization through these hypotheses provides the expected values for the elements of $\vec{V}_v$ for each type (casual, conventional and demonstrative). To use these values in a practical system, we calculate the weighted sum of the elements in $\vec{V}_v$, in which the weights (punctuation) are based on the three assumptions given ($H1$, $H2$ and $H3$).

Thus, each of the $v$ videos has a final score for each hypothesis, where the type of crowd is assigned based on the hypothesis with the highest score. The weights for each element of $\vec{V}_v$ have been defined and are presented in Table 4.1. The symbol – means that the feature does not interfere in that crowd type.

Once the crowd type is defined, all elements of $\vec{V_v}$ are ready to be used in the calculation of similarity. Each video $v$ is represented by the values $H_{1,v}$, $H_{2,v}$ and $H_{3,v}$, as detailed in Table 4.1. The calculation of similarity consists in calculating the distance between two videos of a set of $v$ videos, for that, the distance of *Mahalanobis* is used. In statistics, the distance of *Mahalanobis* is a distance measure introduced by Mahalanobis (1936). It is based on correlations between variables with different patterns.

Then, by considering $\vec{H}_v$ as a vector composed by $(H_{1,v}, H_{2,v}, H_{3,v})$ for a video $v$ and the same for a video $m$ $\vec{H}_m$, the measure of similarity between them is given by:

$$D(\vec{H}_v, \vec{H}_m) = \sqrt{(\vec{H}_v - \vec{H}_m)^T S_v^{-1} (\vec{H}_v - \vec{H}_m)}, \tag{4.11}$$

where $S_v$ is the co-variance matrix of the set $v$ and $D$ is the *Mahalanobis* distance.

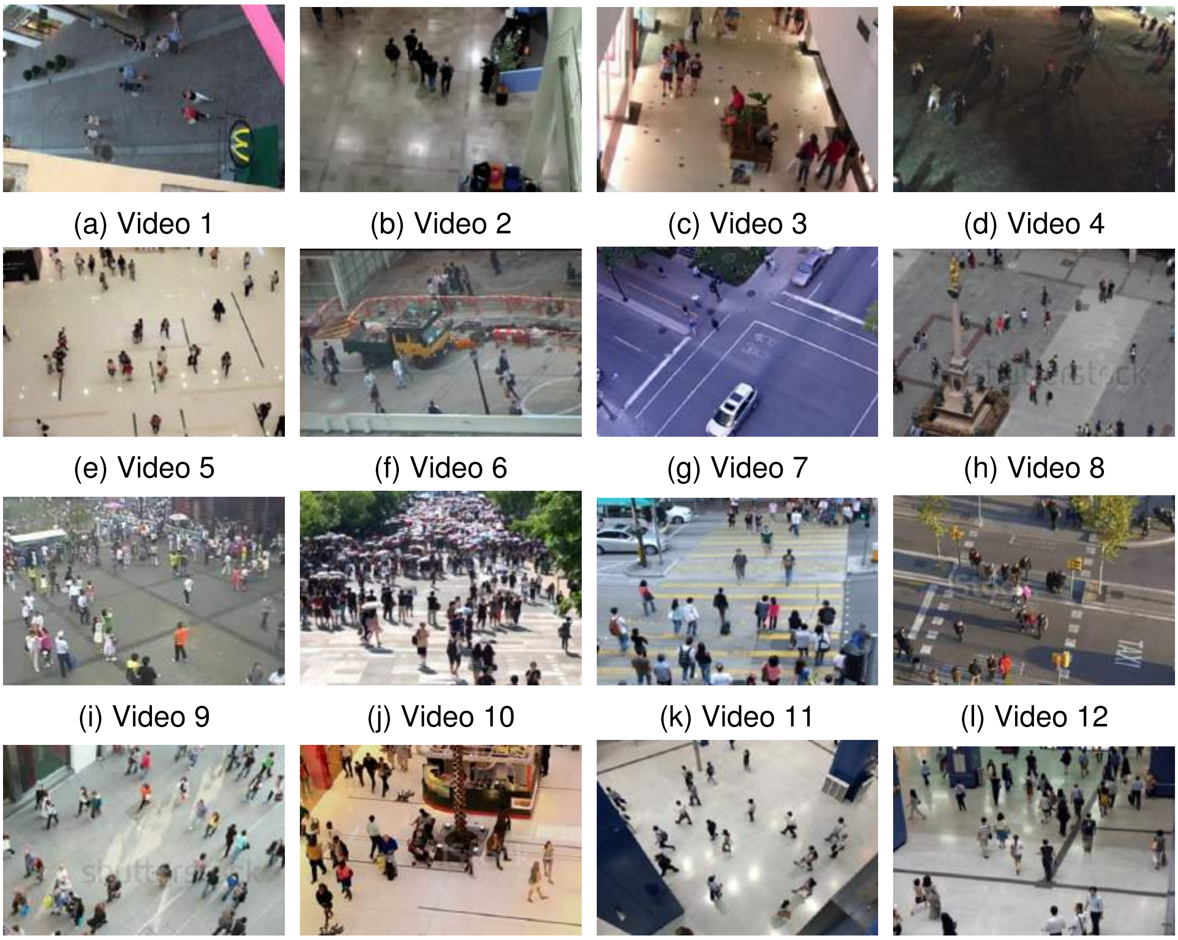
Table 4.1: **Crowd type hypotheses**: calculation for the score of each hypothesis according to each element from the features vector of certain video $v$.

| Feature | *Cluster* | **H1** (Casual) | **H2** (Conventional) | **H3** (Demonstrative) |
|---|---|---|---|---|
| Mean density $Avg\langle\Phi\rangle$ | $S_0$ | 0,2 | 0 | 0 |
| | $S_1$ | 0,1 | 0,1 | 0,1 |
| | $S_2$ | 0 | 0,2 | 0,2 |
| Mean speed $Avg\langle\Theta\rangle$ | $S_0$ | 0 | – | 0 |
| | $S_1$ | 0,1 | – | 0,1 |
| | $S_2$ | 0,2 | – | 0,2 |
| Speed variation $Std\langle\Theta\rangle$ | $S_0$ | – | 0,2 | – |
| | $S_1$ | – | 0,1 | – |
| | $S_2$ | – | 0 | – |
| Mean colletivity $Avg\langle\Psi\rangle$ | $S_0$ | 0,2 | 0 | 0 |
| | $S_1$ | 0,1 | 0,1 | 0,1 |
| | $S_2$ | 0 | 0,2 | 0,2 |
| Grouped pedestrians $\nu$ | $S_0$ | 0 | – | 0 |
| | $S_1$ | 0,1 | – | 0,1 |
| | $S_2$ | 0,2 | – | 0,2 |
| Pedestrians alone $\xi$ | $S_0$ | – | 0 | – |
| | $S_1$ | – | 0,1 | – |
| | $S_2$ | – | 0,2 | – |
| Frequency of groups $\tau$ | $S_0$ | 0,2 | 0 | 0 |
| | $S_1$ | 0,1 | 0,1 | 0,1 |
| | $S_2$ | 0 | 0,2 | 0,2 |

## 4.5 Experimental results

We evaluated our technique running some experiments. Initially, we performed a survey to assess the people understanding as a function of visual video information i.e. data of $\vec{H}_v$. We want to find out if numerical measured data, e.g. size of groups of people, can be perceived in short sequences.

The survey was composed by 16 videos and the subjects were invited to answer five questions about each video. The videos illustrated people walking or standing in several situations. We used videos from different countries obtained from various public databases available on internet (Zhou et al., 2014; Shaikh et al., 2016; Rodriguez et al., 2011). Some representative frames of the dataset used in this analysis are showed in Figure 4.7.



(a) Video 1 (b) Video 2 (c) Video 3 (d) Video 4
(e) Video 5 (f) Video 6 (g) Video 7 (h) Video 8
(i) Video 9 (j) Video 10 (k) Video 11 (l) Video 12
(m) Video 13 (n) Video 14 (o) Video 15 (p) Video 16

Figure 4.7: **Video dataset**: representative frames from all videos used in our analysis about crowd types and video similarity.

In the next sections we show some results obtained in such analysis.

### 4.5.1 Crowd Type Analysis

In order to evaluate the results obtained with the method of classification of the crowds (casual, conventional and demonstrative) an opinion survey was carried out, with the goal of comparing the results of the proposed method with the people's understanding (visual inspection) about the video information. The idea is to find out whether data measured numerically, such as group size, can be perceived by people in short video sequences.

This research was carried out with 16 videos from different countries in different situations. Participants were asked to answer five questions about each video. Before asking the subjects, the concepts about the types of crowds (presented in Section 2.2.1) were presented to assist them in the responses, as briefly presented as follows:

- **Casual crowds** (Kendall, 2016): are relatively large gatherings of people who happen to be in the same place at the same time; if they interact at all, it is only briefly. People in a shopping mall or a subway car are examples of casual crowds. Other than sharing a momentary interest, such as a clown's performance or a small child's fall, a casual crowd has nothing in common;

- **Conventional crowds** (Kendall, 2016): are made up of people who come together for a scheduled event and thus share a common focus. Examples include religious services, graduation ceremonies, concerts, and college lectures. Each of these events has pre-established schedules and norms. Because these events occur regularly, interaction among participants is much more likely; People leaving events or environments can also be examples;

- **Demonstrator crowds** (Berlonghi, 1995): are crowds who often have a recognized leader, organized for a specific reason or event, to picket, demonstrate, march, or chant.

The (multiple choice) questions with the possible answers used in this survey are described in Table 4.2.

The survey was answered by 10 people resulting in 800 responses (80 answers from each subject). The results were used to validate our approach of group definition and group features. Since each data in our method was clustered in 3 levels, we mapped the answers to the possible levels. We considered correct when the major part of subjects answer in accordance with the higher hypothesis. Figure 4.8 shows the correctness of the method in comparison with the people answers in the survey.

Regarding *Q1*, although the correctness rate is not bad (88%), some videos were misclassified, in our opinion, by the subjects. Indeed, the classification is not

Table 4.2: **Survey on crowds classification**: questions and possible answers used in the research.

| Question | Possible answers |
|---|---|
| **Q1**: In your opinion, which of the following best describes the crowd type in the video? | *a*) Casual crowd;<br>*b*) Conventional crowd;<br>*c*) Demonstrative crowd;<br>*d*) None of them;<br>*e*) I don't know. |
| **Q2**: About groups (people walking together, to the same direction, with similar velocities), do you think the major part of people are grouped or alone in this video? | *a*) Grouped;<br>*b*) Alone;<br>*c*) I don't know. |
| **Q3**: About the size of the groups (quantity of people by group), which of the following you might noticed? | *a*) Small groups ($G < 3$ people);<br>*b*) Big groups ($G \geq 3$ people);<br>*c*) There are no groups;<br>*d*) I don't know. |
| **Q4**: About the crowd density (amount of people by square meter), which of the following you might noticed? | *a*) Low density;<br>*b*) Medium density;<br>*c*) High density;<br>*d*) I don't know. |
| **Q5**: In this video, might you noticed any interaction between people? | *a*) No, there are no interactions;<br>*b*) Yes, few interactions;<br>*c*) Yes, many interactions;<br>*d*) I don't know. |

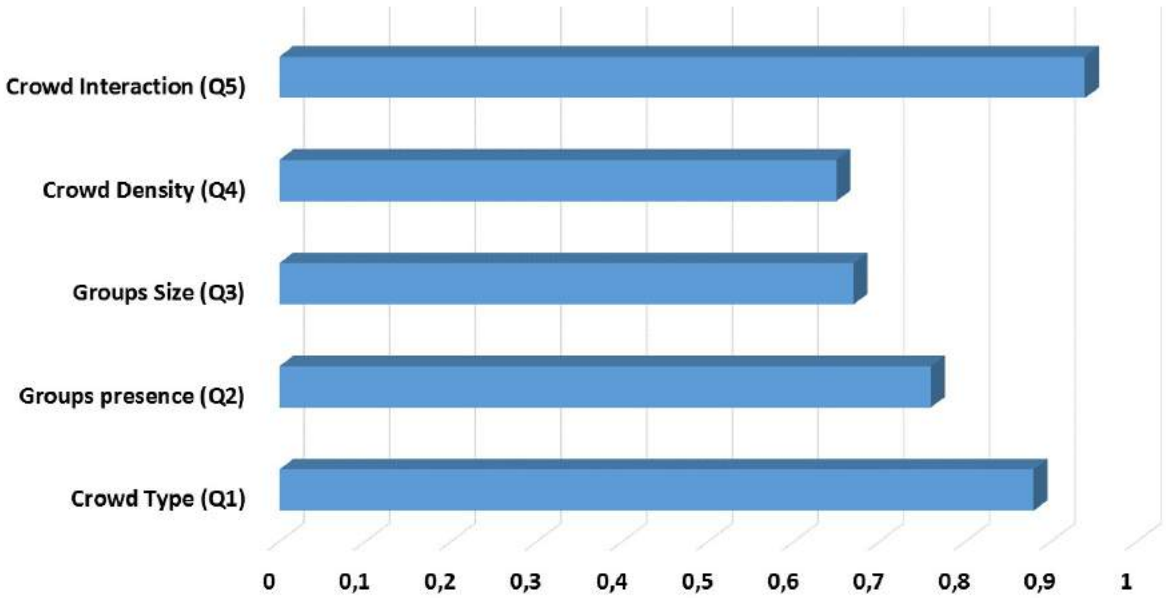



Figure 4.8: **Method correctness**: correctness of our method when compared to the subject answers.

obvious, mainly because the difference between the crowd types (casual vs. conventional) and (conventional vs. Demonstrator) is not very clear in the video sequences.

Concerning *Q2* (76%), related to the presence of groups, the question seems difficult to really measure the reality. For example, 70% of answers chosen option <Alone> for Video 5 (see Figure 4.9), and numerically it is corrected (*Total number of people in the videos:* 21, *Number of Groups:* 4, *Mean people by group:* 2.25 and *Number of non-grouped people:* 12.). Our method selects <Medium> for this metric, which is also correct, but not correspond to the performed question (Grouped or alone?). Although the number of people alone is slightly higher than that of people grouped (12 and 9, respectively), the two figures are quite close, suggesting that there are not many groups, nor many people alone.


Figure 4.9: **Groups in Video 5**: group detection in Video 5, which had the number of groups (four groups, being two permanent, in red, and two temporary, in blue) considered as *Medium* by the proposed method.

Concerning *Q3* (67%), related to the size of the groups, our method presented some errors in groups detection. For instance, in the upper part of the Video 11 (see Figure 4.10(a)), although people are not grouped, they spend a little time near each other due to the probable characteristics of the environment (a pedestrian crossing, with an agglomeration of people when the signal opens). In this case, the method detected groups, but the subjects perceive that people are not actually grouped.

(a) Groups detected in Video 11

(b) Different frames from Video 11

Figure 4.10: **Groups and density from Video 11**: groups detected in Video 11, which had a group mistakenly detected in the upper part due to the nature of the video (a) and several frames of this video, each one of them suggesting a different value for the density (b).

We expected that *Q4* (65%) was easily to visually assess the information about crowd density, however it seems to be a problem of concepts of low and medium densities, mainly when the crowd is medium density in comparison with low and high. For example, considering the Video 11 (where some frames are illustrated in Figure 4.10(c)), it is possible to observe different levels of densities throughout the video, which may have confused subjects to answer this question. The proposed method considers the <*Medium*> density throughout the video, while the subjects are faced with different densities along the frames.

Finally, the proposed method presented the highest correction rates in *Q*5 (94%) question, which analyzes the interactions in the crowd, indicating that participants were able to visually assess the interaction between pedestrians in the video. The results obtained in the video similarity analysis are presented below.

### 4.5.2 Video Similarity Analysis

Similarity analysis among crowds was performed with 33 videos[2] (16 of them are illustrated in Figure 4.7). Using Equation 4.11 (distance of *Mahalanobis*, discussed in Section 2.2.1), all possible combinations ($n^2$) of similarity were calculated, where $n$ is the number of $v$ videos analyzed. Figure 4.11 shows the similarity among each video.

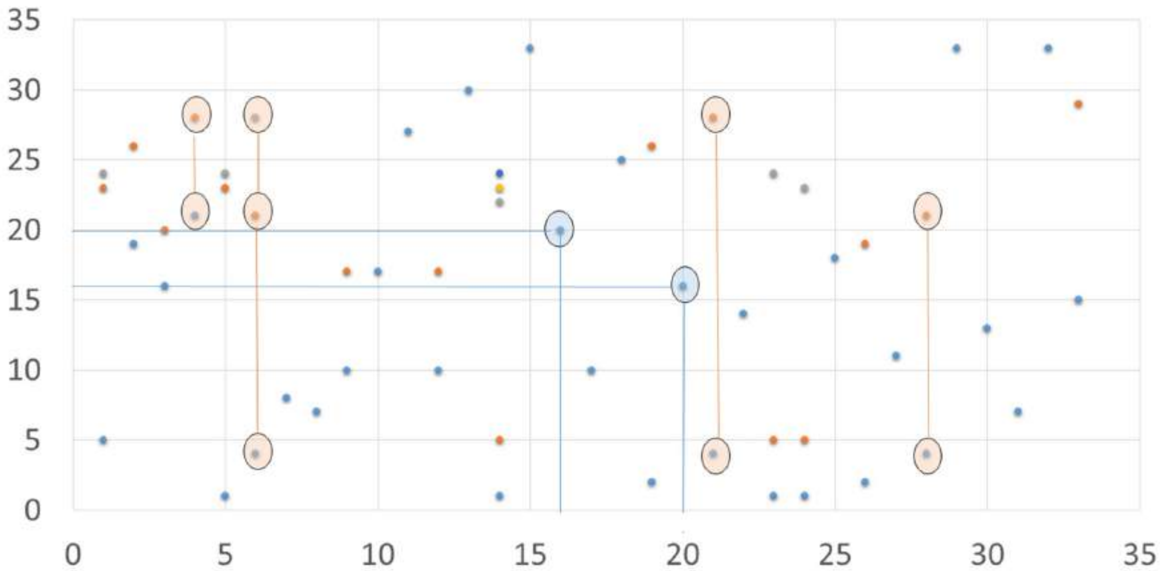



Figure 4.11: **Similarity among videos**: blue ellipse on the left define the most similar crowd with video 16, i.e. video 20. In a reciprocal way, the blue ellipse on the right presents the most similar crowd with video 20, i.e. video 16. Similarly, the most similar crowds in videos 4, 6, 21 and 28 are highlighted with orange ellipses.

Firstly, we present in Figure 4.11 one of pairs which present the smallest distance $D$ (some of them present the same *Malahanobis* distance $D$). Indeed, this graphic is a matrix where rows and columns represent the video index. The plot highlights videos which distances $D$, between the hypothesis values ($\overrightarrow{H_v}$), are smaller.

[2] In addition to the 16 video sequences selected showed in Figure 4.7, other videos from different countries and three others of unknown nationalities were used. Since the goal was to compare crowds and, with no concerns about cultures, knowing the country of origin of each crowd was not a prerequisite.

For example, the videos $v$ = 16 and $v$ = 20 (highlighted with blue ellipses in Figure 4.11) are reciprocally the most similar to each other, according to the proposed metric. Indeed, the left blue ellipse on Figure 4.11 represents the most similar crowd with video $v$ = 16, i.e. video $v$ = 20. In a reciprocal way, the blue ellipse on the right presents the most similar crowd with video $v$ = 20 i.e. video $v$ = 16. To provide a qualitative evaluation of this result, Figure 4.12 illustrates the frames of both videos, which appear to be similar.


(a) Video $v$ = 16 (b) Video $v$ = 20

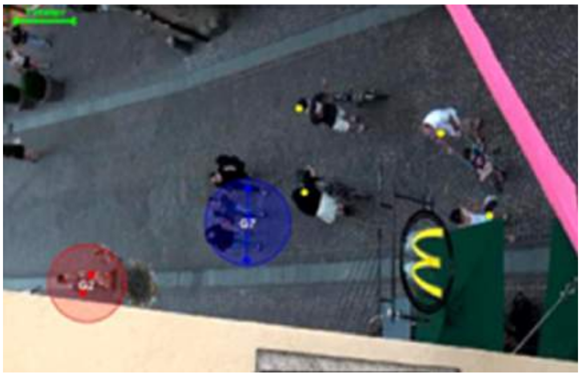
Figure 4.12: **Videos $v$ = 16 and 20**: red circles represent permanent groups and blue circles represent temporary groups. Yellow dots indicate ungrouped pedestrians.

In addition to this qualitative analysis, the similarity between these two videos is repeated in numerical data, as described in Table 4.3.

Table 4.3: **Videos $v$ = 16 and 20**: information about similarity.

| | Video $v$ = 16 | Video $v$ = 20 |
|---|---|---|
| Total number of pedestrians | 26 | 28 |
| Number of Groups | 6 | 6 |
| Mean pedestrians by group | 2.33 | 2.55 |
| Number of non-grouped pedestrians | 12 | 13 |

Similarly, the crowds of videos $v$ = 4, $v$ = 6, $v$ = 21 and $v$ = 28 also have presented short distances $D$ (*Malahanobis*) among them or, in other words, crowds are similar according to the proposed method (these videos are highlighted with orange ellipses in Figure 4.11). Figure 4.13 illustrates the most representative frames of these videos, where it is possible to observe the visual similarity between them and the Table 4.4 shows the numerical data.




(a) Video $v$ = 4

(b) Video $v$ = 6


(c) Video $v$ = 21


(d) Video $v$ = 28

Figure 4.13: **Videos** $v$ = **4, 6, 21 and 28**: red circles represent permanent groups and blue circles represent temporary groups. Yellow dots indicate ungrouped pedestrians.

## 4.6 Chapter Remarks

This chapter presented and discussed the results obtained in the initial steps of the geometrical dimensions model proposed in this thesis. Basically, after a tracking and planar homography process, some features are extracted from each pedestrian. These features characterizes the *I - Physical* and *II - Social* dimensions, illustrated in Figure 1.1 (displayed in the Introduction chapter). These features are: angular variation, cohesion, collectivity, density, distance, speed, groups, ungrouped pedestrians and number of pedestrians (number of pedestrians is a crowd feature, while the others are related to each pedestrian). Socialization and Isolation features from *II - Social* dimension are described in the next chapter. Table 4.5 shows the information about the occurrence of these features as a function of the group and pedestrian.

Table 4.4: **Videos** $v$ = **4, 6, 21 and 28**: information about similarity.

| | $v = 4$ | $v = 6$ | $v = 21$ | $v = 28$ |
|---|---|---|---|---|
| Total number of pedestrians | 9 | 10 | 13 | 6 |
| Number of Groups | 2 | 2 | 2 | 1 |
| Mean pedestrians by group | 2.5 | 2.5 | 2 | 2 |
| Number of non-grouped pedestrians | 4 | 5 | 9 | 4 |

Table 4.5: **Features occurrence**: information about each feature occurrence: (•) indicate the feature is calculated in that level.

| | Feature | At each frame | By pedestrian | By group |
|---|---|---|---|---|
| *I - Physical* | Angular Variation | • | • | • |
| | Density | • | | |
| | Distance | • | | • |
| | Speed | • | • | • |
| *II - Social* | Area | • | | • |
| | Cohesion | • | | • |
| | Collectivity | • | • | |
| | Isolation | • | • | |
| | Socialization | • | • | |

The purpose of Table 4.5 is to show whether each of the characteristics is calculated for each pedestrian, for each group or both (pedestrians and groups). For example, cohesion (meaning how much pedestrians want to stay in the group) is a specific feature of a group of pedestrians, while isolation is a specific feature of pedestrians.

The features that belong to pedestrians and groups at the same time (such as speed and angular variation), in the case of the group level, have values obtained through the arithmetic mean of the values of these features for the pedestrians who are part of the group.

In addition, we also presented a metric to find out similarity between crowds in videos and the detection of the crowd type. It is important because when comparing two crowds of different countries to find cultural differences, it is necessary that they have similar characteristics, for example, it would not be correct to compare a crowd with many hurried people at a train station in China with a crowd of people strolling leisurely in a park in Italy, there would probably be differences masked by the features of the places where the videos were recorded. This results are published in (Favaretto et al., 2016b). In our dataset, presented in details in Appendix A (*Video Analysis Dataset and Applications*), we included this similarity analysis.

In the next chapter (*Third Dimension: Detection of Personality and Emotion Traits*), we described how we use the features presented is this chapter to map it into features from the *III - Personal and Emotional* dimension.

# 5. THIRD DIMENSION: DETECTION OF PERSONALITY AND EMOTION TRAITS

This chapter presents the proposed methodology to detect personality and basic emotion characteristics of crowds in video sequences, features from the *III - Personal and Emotional* dimension. Firstly, individuals are detected and tracked, then such information is mapped to OCEAN dimensions, used to find out personality and emotion in videos, based on OCC emotion model.

Section 5.1 presents the proposed approach to extract data from pedestrians in the video. Section 5.2 describe how the extracted data is mapped into personality (OCEAN) and emotion (OCC) traits. Section 5.3 discuss some results and analysis performed in spontaneous videos, comparing the results with the psychology literature in this area. Finally, Section 5.4 shows an analysis about the perception of personality and emotion features by the users in videos of pedestrians.

## 5.1 Proposed Approach for data Extraction

In this chapter, we propose to detect personality traits based on the Big-five personality model (presented in details in Section 2.1.2 of this manuscript), also referred as (OCEAN), using individuals behaviors automatically detected in video sequences. For this, we used the NEO PI-R (Costa and McCrae, 2007) which is the standard questionnaire measure of the Five Factor Model.

In addition, we use such psychological traits to identify some primordial emotions in the crowd. Using a similar mapping as proposed by Saifi et al. (2016), we were able to identify the level of some emotions for each individual, like happiness or fear, according to its OCEAN level. Figure 5.1 shows an overview of the proposed method.

Our model presents three main steps as following: video data extraction, personality and emotion analysis. These steps are illustrated in the overview of the method in Figure 5.1. The first step aims to obtain the individual trajectories from observed pedestrians in real videos, as described in Section 4.1, in the previous chapter. Using these trajectories, we detect groups and extract data which are useful for second step, that is responsible for personality mapping of pedestrians,

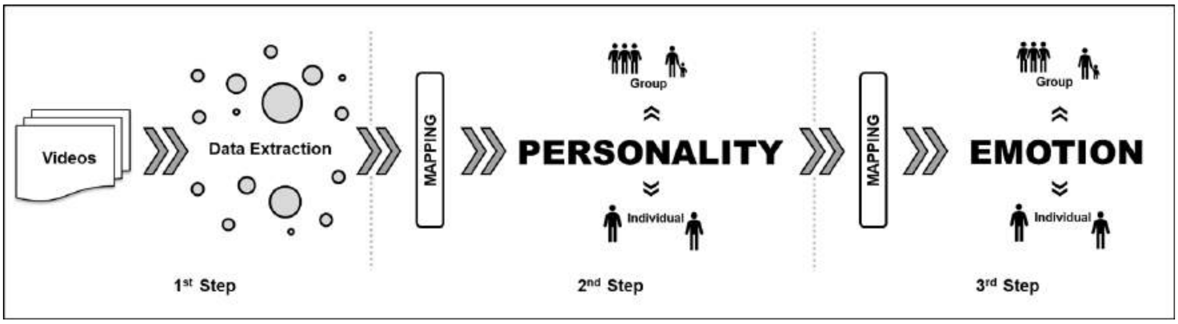



Figure 5.1: **Main steps of personality and emotion detection**: ($1^{st}$) individual tracking and data extraction, ($2^{nd}$) individual data is mapped to individual and group personality (OCEAN) traits and ($3^{rd}$) personality traits are mapped to individual and group basic emotions. These steps are detailed in Figure 5.2.

as described in Section 5.2. Once we have concluded the second step, we have enough information to follow with the third step, which consists of emotion detection of individuals and groups according to OCEAN values. These steps are presented in Section 5.2. Figure 5.2 illustrates a more detailed flow chart of our approach.

### 5.1.1 Individual Data Extraction

Initially, the information about pedestrians is obtained as described in Section 4.1 (*Initial Detection and Tracking*) in the previous chapter. Then we use the planar homography to rectify the image perspective, as presented in Section 4.1.1 in this manuscript. Based on the pedestrian trajectories, we compute, in addition to the geometric information for each pedestrian $i$ described in previous chapter (2D position $x_i$, speed $s_i$, angular variation $\alpha_i$ and collectivity $\phi_i$), two other features: socialization $\vartheta_i$ and isolation levels $\varphi_i$.

These features were chosen because two reasons: Firstly, they are strongly related with the questions concerned with groups activities in Neo-Pi survey (Costa and McCrae, 2007). The second reason is the theory behind socialization/isolation that easily can be represented through geometric data (positions and distances), and collectivity that has been already explored in the context of crowd behaviors detection (Zhou et al., 2014).

To compute the socialization level $\vartheta$ we use a classical supervised learning algorithm proposed by Moller Moller (1993). The artificial neural network (ANN)

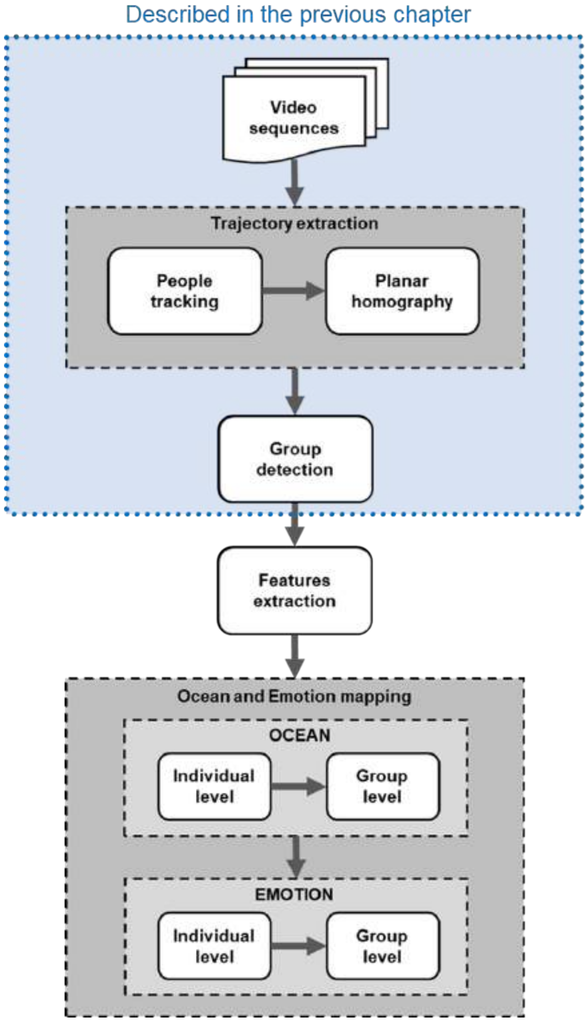



Figure 5.2: **Flow chart of our approach**: initially, pedestrian trajectories are extracted and passed by a homography transformation process. Based on these trajectories, groups are detected and its features are extracted. These features are the inputs for the OCEAN and EMOTION acquisition process.

(illustrated in Figure 5.3) uses a Scaled Conjugate Gradient (SCG) algorithm in the training process to calculate the socialization $\vartheta_i$ level for each individual $i$.

As described in Figure 5.3, the ANN has 3 inputs (collectivity $\phi_i$ of person $i$, mean Euclidean distance from a person $i$ to others $\bar{d}_{i,j}$ and the number of people in the Social Space[1] according to Hall's proxemics (Hall, 1966) around the person $n_i$).

In addition, the network has 10 hidden layers and 2 outputs (the probability of socialization and the probability of non socialization). The final accuracy from the training processes was 96%. We used 16.000 samples (70% of training and 30%

[1]Social space is related to 3.6 meters according to Hall (1966).

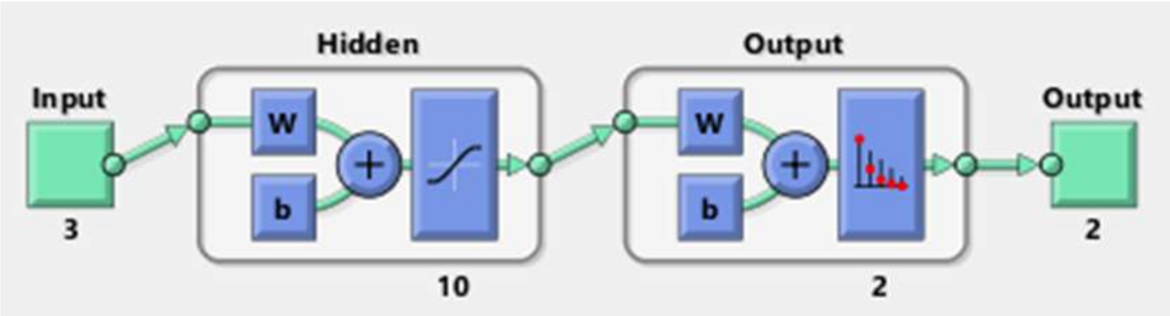



Figure 5.3: **ANN for socialization**: neural network used to learn the socialization level.

of validating). These samples were obtained from the 25 initial frames from each of the videos from our dataset. The remaining frames were used to test the ANN.

The ground truth (GT) was generated as follows: Firstly, we define if a person has a high socialization level $GT\_\vartheta_i$ based on Hall's *proxemics*, calculated according to the Equation 5.1:

$$GT\_\vartheta_i = \begin{cases} 0, & \text{if } n_i = 0; \\ \frac{n_i}{\rho}, & \text{otherwise.} \end{cases} \quad (5.1)$$

where $n_i$ is the number of individuals in the social space around the person $i$ and $\rho$ is the number of individuals in the analyzed frame. If $GT\_\vartheta_i >= 0.5$, we considered this person as a "social" person, otherwise the person is considered "not social" in the training processes. Secondly, we proceed a visual inspection manually correcting false positives or false negatives in comparison to our personal opinion. Using this GT and the neural network, we evaluate $\vartheta_i$ for each individual $i$ at each frame, for each video in the test group. Once we get the socialization level $\vartheta_i$, we compute the isolation level $\varphi_i = 1 - \vartheta_i$, that corresponds to its inverse.

Finally, for each pedestrian $i$ in a frame $f$ of a certain video $v$, we will have a features vector $\overrightarrow{V_{i,v}^f} = \left[x_{i,v}^f, s_{i,v}^f, \alpha_{i,v}^f, \phi_{i,v}^f, \vartheta_{i,v}^f, \varphi_{i,v}^f\right]$. Then, when we compute the average for an individual $i$, for all frames of a video $v$, we will have a vector $\overrightarrow{V_{i,v}}$ for each pedestrian $i$.

Next section show how the vector $\overrightarrow{V_{i,v}}$ of features is mapped into personalities and emotions.

## 5.2 Features Mapping

In this section is described how we proceed to map the features vector from each individual in a specific video $\overrightarrow{V_{i,v}}$ to OCEAN dimensions (Section 5.2.1). Once we get the OCEAN values, we proceed to emotion mapping (Section 5.2.2).

### 5.2.1 Big-Five Mapping

Our goal is to map data from $\overrightarrow{V_i}$ to $\overrightarrow{BF_i}$, where the last one is related to the Big-Five dimensions (or OCEAN) for each individual $i$ for a certain video and described as a features vector: $\overrightarrow{BF_i} = [O_i, C_i, E_i, A_i, N_i]$.

We used the empirically equations[2] to map individual and group characteristics in OCEAN cultural dimensions. Basically, the method proposes to "answer" 25 items from NEO PI-R inventory for each individual in the video sequence. All possible answers for each NEO PI-R item are in the interval $[1;5]$ (indicating, respectively, Strongly Disagree, Disagree, Neutral, Agree and Strongly Agree), being 1 related to Strongly Disagree and 5 to Strongly Agree for a referred question.

Although the complete version of NEO PI-R has 240 items, 25 of them were selected since they have a direct relationship with crowd behavior. Each question is mapped to an equation using only information contained in $V$ for each $i$. For example, in order to represent the item "1. Have clear goals, work to them in orderly way", we consider that the individual $i$ should have a high velocity $s$ and low angular variation $\alpha$ to have answer compatible with "Strong". So the equation for this item is $Q_1 = s_i + \frac{1}{\alpha_i}$. Table 5.1 shows the equations used for each considered question.

Once all questions $k$ (in the interval $[1;25]$) have been answered for all individuals $i$, we have $\overrightarrow{Q^f_{i,k}}$ for each frame $f$. In addition, we computed the average values to have one vector $\overrightarrow{Q_{i,k}}$ per video. According to NEO PI-R definition, each of the questions $\overrightarrow{Q_{i,k}}$ is associated to one of the Big Five dimensions and some questions should invert the values, because an item score 4 (Strongly Agree) can represent a high or low value of a certain personality trait. So, to get the correct values, we applied a factor to the questions resulting in $\overrightarrow{Q^*_{i,k}}$ or $\overrightarrow{Q'_{i,k}} = 4 - \overrightarrow{Q_{i,k}}$,

[2] These equations were proposed based on Psychology literature and supervised and evaluated by two collaborators of this research: Angelo Brandelli Costa and Felipe Vilanova, both from the Department of Psychology, PUCRS.

Table 5.1: **NEO PI-R equations**: equations from each NEO PI-R selected item.

| NEO PI-R Item | Equation |
|---|---|
| 1. Have clear goals, work to them in orderly way | $Q_1 = s_i + \frac{1}{\alpha_i}$ |
| 2. Follow same route when go somewhere | $Q_2 = \alpha_i$ |
| 3. Shy away from crowds<br>4. Don't get much pleasure chatting with people<br>5. Usually prefer to do things alone<br>6. Prefer jobs that let me work alone, unbothered<br>7. Wouldn't enjoy holiday in Las Vegas<br>8. Many think of me as somewhat cold, distant | $Q_{3-8} = \varphi_i$ |
| 9. Rather cooperate with others than compete<br>10. Try to be courteous to everyone I meet | $Q_{9-10} = \phi_i$ |
| 11. Social gatherings usually bore me | $Q_{11} = \varphi_i + std(\alpha_i)$ |
| 12. Usually seem in hurry | $Q_{12} = s_i + \alpha_i$ |
| 13. Often disgusted with people I have to deal with | $Q_{13} = \varphi_i + \frac{1}{\phi_i}$ |
| 14. Have often been leader of groups belonged to | $Q_{14} = \phi_i + \vartheta_i + \frac{1}{\alpha_i}$ |
| 15. Would rather go my own way than be a leader | $Q_{15} = \frac{1}{Q_{14}}$ |
| 16. Like to have lots of people around me<br>17. Enjoy parties with lots of people<br>18. Like being part of crowd at sporting events<br>19. Would rather a popular beach than isolated cabin<br>20. Really enjoy talking to people<br>21. Like to be where action is | $Q_{16-21} = \vartheta_i$ |
| 22. Feel need for other people if by myself for long<br>23. Find it easy to smile, be outgoing with strangers<br>24. Rarely feel lonely or blue<br>25. Seldom feel self-conscious around people | $Q_{22-25} = \vartheta_i + \phi_i$ |

depending on they must be inverted or not. The final equations of each OCEAN dimension are:

$$O_i = \frac{Q^*_{i,2}}{\varrho}, \tag{5.2}$$

$$C_i = \frac{Q'_{i,1}}{\varrho}, \tag{5.3}$$

$$E'_i = Q'_{i,3} + Q'_{i,12} + Q'_{i,14} + \sum_{q=16}^{23} Q'_{i,q}, \tag{5.4}$$

$$E^*_i = \sum_{q=4}^{8} Q^*_{i,q} + Q^*_{i,11} + Q^*_{i,15}, \tag{5.5}$$

$$E_i = \frac{(E'_i + E^*_i)}{\varrho}, \tag{5.6}$$

$$A_i = \frac{\sum_{q=9}^{10} Q'_{i,q}}{\varrho}, \tag{5.7}$$

$$N_i = \frac{Q'_{i,13} + \sum_{q=24}^{25} Q^*_{i,q}}{\varrho}, \tag{5.8}$$

where $\varrho$ represents the percentage of questions from the total, in each dimension (O, C, E, A and N), respectively 4%, 4%, 72%, 8% and 12%. In the next section is presented the approach to map the personality traits in emotions.

### 5.2.2 Emotion Mapping

In this section we present how we connected personality and emotion traits in order to detect data in the individuals from video sequences. As mentioned by Revelle and Scherer (2009): “A helpful analogy is to consider that personality is to emotion as climate is to weather. That is, what one expects is personality, what one observes at any particular moment is emotion”. That is the reason why, in our work, we correlated emotion traits with visual behaviors that will be perceived in pedestrian motion.

Our work proposes an emotion mapping based on personality traits (i.e. OCEAN) found for each individual present in the video sequence. First, we selected four emotions from OCC (Ortony et al., 1990) model: Fear, Happiness, Sadness and Anger. It is important to notice that we chose only four emotions that, in our opinion, are the most visible when relating with motion behavior, which is our case in video sequences with pedestrians. Any other from the total of 22 emotions proposed in OCC model could also be mapped.

Once we have computed the personality traits for an individual, we propose a way to map these personality traits into the four considered emotions. In addition to vector $\vec{V}_i^f$ for individual *i* per frame *f* described in Section 5.1, we included personality and emotion parameters as follows: $\vec{V}_i^f = \left[s_i^f, \alpha_i^f, \varphi_i^f, \vartheta_i^f, \phi_i^f, \vec{P}_i^f, \vec{E}_i^f\right]$. $\vec{P}_i^f$ states for values of OCEAN personality $\left[O_i^f, C_i^f, E_i^f, A_i^f, N_i^f,\right]$, while $\vec{E}_i^f$ states for values of emotion based on OCC model: $\left[F_i^f, H_i^f, S_i^f, An_i^f\right]$. Values of $\vec{P}$ are in the interval [0; 1] while values of $\vec{E}$ are from $[-3; 3]$. For groups analysis, we only considered parameters of personality and emotion in the detected groups $g$ in the video sequence. In addition to previously defined $\vec{G_g} = \left[n_g, \vec{I}_g, \bar{s}_g, \bar{\alpha}_g, \bar{d}_g\right]$, we include $\vec{P}_g$ and $\vec{E}_g$ containing the average values of $n_g$ individuals in $g$.

In order to map from OCEAN to emotion parameters we observe some aspects in literature, as the ones proposed by Costa and McCrae (2007):

- O- : person is close to interact with others;
- O+ : person is aware of his/her feelings;
- C+ : person is optimistic;
- C- : person is pessimist;
- E+ : person has a strong relationship with positive emotions;
- E- : person presents relationship with negative emotions;
- A+ : person has a strong relationship with positive reactions;
- A- : person presents relationship with negative reactions;
- N-: known by the emotional stability;
- N+ : person feels negative emotions.

Such data resulted in empirical definitions included in Table 5.2, that shows the mapping from OCEAN traits to the chosen emotions. In fact, we were not the first one to propose this type of mapping. Saifi et al. (2016) proposes similar data for a different approach where authors were interested in providing critical emotions in crowd simulators. Davis and Panksepp (2011) also proposed a similar approach unifying basic emotions with personality.

In Table 5.2, the plus/minus signals along each factor represent the positive or negative values of each one. For example, O+ stands for positive values (i.e. O $\geq 0.5$) and O- stands for negative values (i.e. O $< 0.5$)). A positive value for a given factor (i.e. 1) means the stronger the OCEAN trait is, the stronger is the emotion too. A negative value (i.e. -1) does the opposite, therefore, the stronger the factor's value, the weaker is a given emotion. A zero value means that a given emotion is not affected at all by the given factor. To better illustrate, a hypothetical example is given: if an individual has a high value for Extraversion (for example, E = 0.9), following the mapping in Table 5.2, this individual can present signals of happiness (i.e. If E+ then Happiness= 1) and should not be angry (i.e. If E+ then Anger= -1).

Table 5.2: **Emotion mapping from OCEAN to OCC**: the plus/minus signals along each factor represent the positive/negative value of each one.

| **OCEAN Factors** | **Fear** *F* | **Happiness** *H* | **Sadness** *S* | **Anger** *An* |
|---|---|---|---|---|
| O+ | 0 | 0 | 0 | -1 |
| O- | 0 | 0 | 0 | 1 |
| C+ | -1 | 0 | 0 | 0 |
| C- | 1 | 0 | 0 | 0 |
| E+ | -1 | 1 | -1 | -1 |
| E- | 1 | 0 | 0 | 0 |
| A+ | 0 | 0 | 0 | -1 |
| A- | 0 | 0 | 0 | 1 |
| N+ | 1 | -1 | 1 | 1 |
| N- | -1 | 1 | -1 | -1 |

In the next section we present some results and discussions regarding OCEAN and emotion detection.

## 5.3 Results and Discussions

In this section we discuss some results obtained with this dimension of our approach. We organized it into two different analysis: *i)* Ocean and Emotion recognition in spontaneous videos (Sections 5.3.1 and 5.3.2, respectively) and *ii)* an Ocean and Emotion comparison with Literature (Section 5.3.3).

### 5.3.1 OCEAN Analysis

In this section, we present OCEAN analysis involving spontaneous videos (crowds in public spaces). Firstly, we calculate the OCEAN of each individual in the video at each frame. Once we get the individuals OCEAN, the group OCEAN is computed by the average of the individuals' OCEAN that are part of the group. Figure 5.4 shows a representation of each individual OCEAN in a determined frame. We used five color box that represent the five dimensions, where blue is related to Openness, cyan indicates Conscientiousness, green indicates Extraversion, yellow means Agreeableness and red, Neuroticism.

The group of individuals highlighted in Figure 5.4 is composed of two people. The OCEAN of this group (named $G1$) over the time is illustrated in Figure 5.5(a), and obtained by the average OCEAN of its individuals ($P9$ and $P10$), presented in Figures 5.5 (b) and (c). As can be seen in such figures, the two individuals present more motion variation at the beginning of the video and then they keep the same motion characteristics until the end of the short movie (which duration is 100 frames). Their Openness is high because they keep low angular variation along their trajectories. On the other hand, their Conscientiousness dimension is lower than other dimensions because they keep low speeds in comparison to other groups.

The group OCEAN reflects the individuals OCEAN. An analysis of OCEAN in the country level is presented later in this section. Once we have the OCEAN values, individuals emotions are detected in the analyzed videos, as discussed in next section.

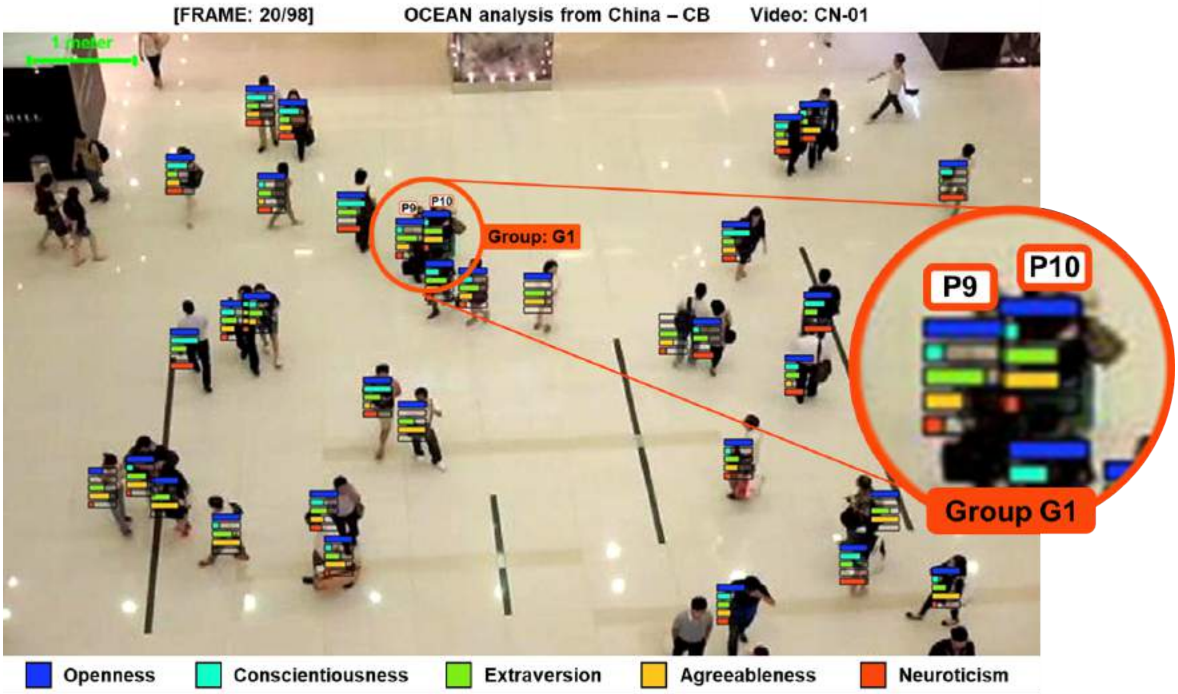


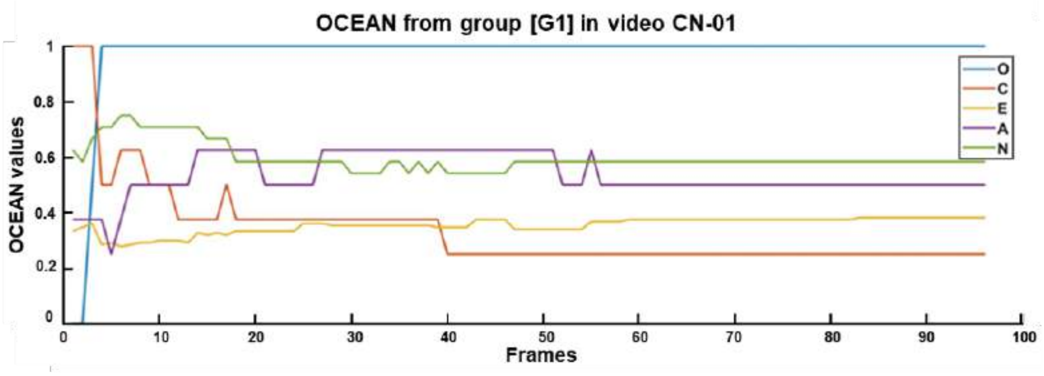

Figure 5.4: **OCEAN representation of each individual in the scene**: each color bar is related to one OCEAN dimension: blue is *Openness*, cyan is *Conscientiousness*, green is *Extraversion*, yellow is Agreeableness and red, *Neuroticism*. Information about highlighted group $G1$ (formed by the individuals $P9$ and $P10$) is detailed in Figure 5.5.

### 5.3.2 Emotion Analysis

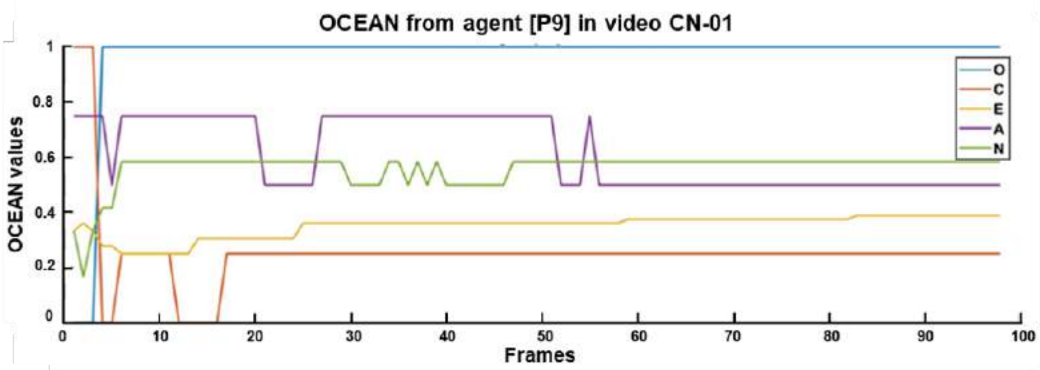
This section presents the emotion analysis in spontaneous videos. Figure 5.6 shows an example of the emotion detection in a video from Austria. A filled square represents that the person has a positive value from that emotion, a half filled square means that the emotion is neutral and a not filled square means that the person has a negative value from the emotion.

For example, the highlighted person, with the blue arrow in Figure 5.6, got a negative value regarding the Anger emotion (the red square is not filled). It happens because this individual is interacting with the other (walking in the same direction and with similar angular variation), so collectivity and socialization levels are high and isolation level is low, consequently Anger receives a negative score.

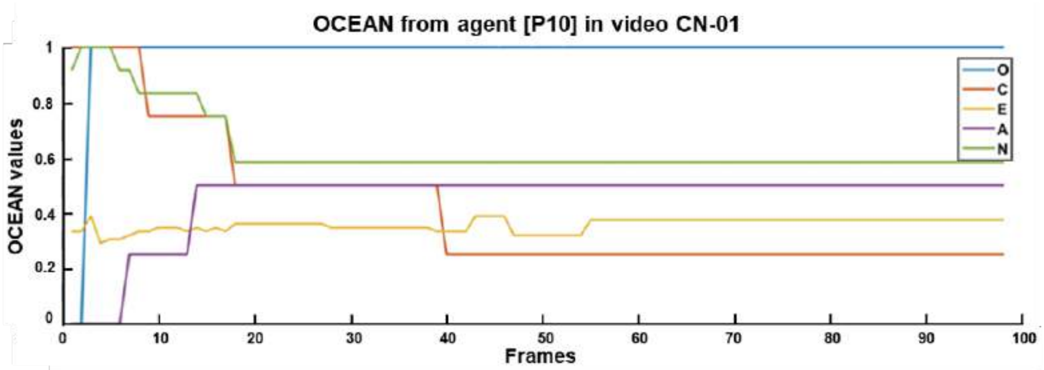
Still in Figure 5.6, the highlighted person with the orange arrow gets a zero value for the Happiness emotion (the green square is half filled). It happens because she/he is alone, changing orientation (angular variation), with high isolation. The




(a) Group $G1$




(b) Individual $P9$




(c) Individual $P10$

Figure 5.5: **OCEAN observed in the group** $G1$: OCEAN data over time observed in the group of individuals highlighted in Figure 5.4: data from group $G1$ (a) and its individuals $P9$ (b) and $P10$ (c).

person highlighted with the red arrow gets a positive value for the Fear (the yellow square is completely filled). It happens because this individual is moving slower and with high angular variation, so his/her dimension C is low, generating a high value for Fear. The legend of colors is the following: red is related to Anger, yellow means Fear, green indicates Happiness and blue, Sadness.

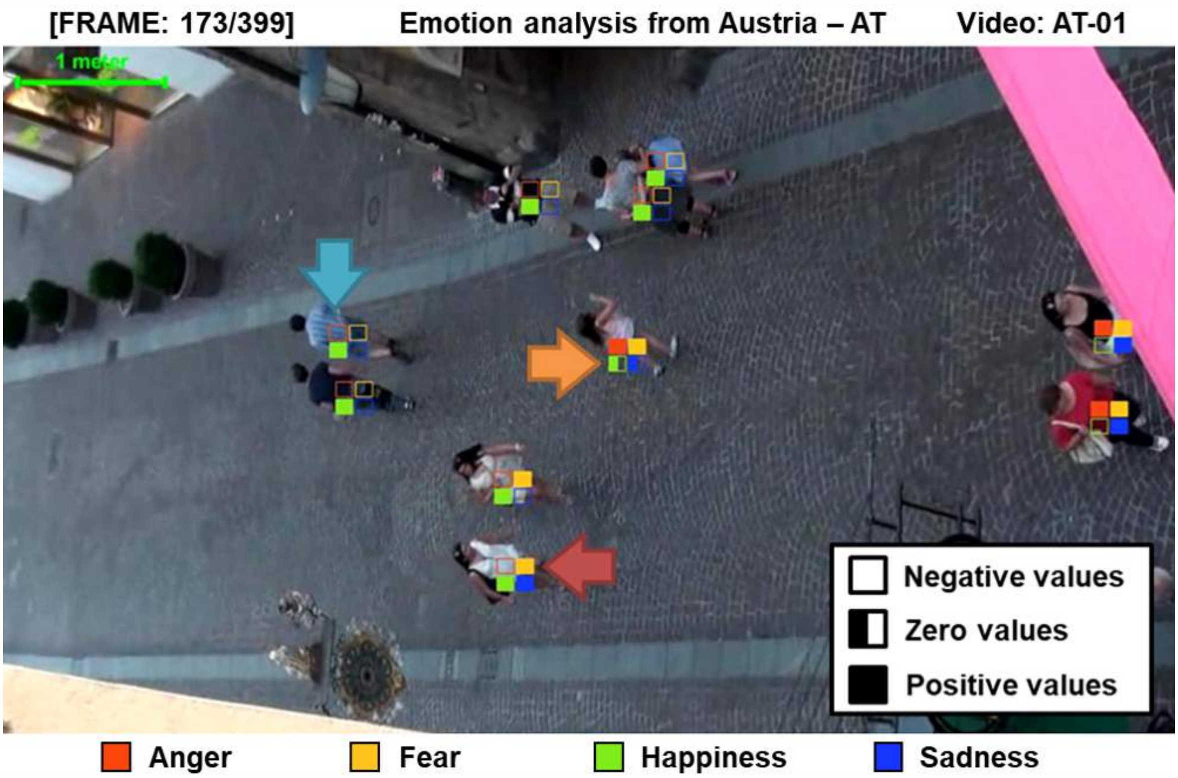
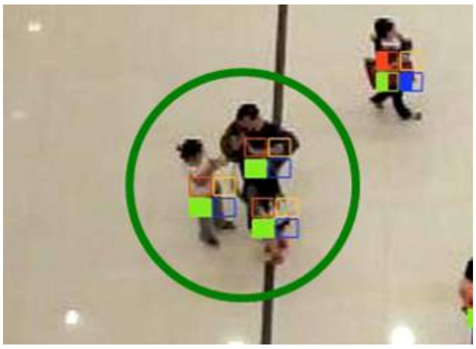



Figure 5.6: **Emotion representation based on the OCEAN mapping process**: each square represents one emotion, where a filled square represents that the person has a positive value from an emotion, a half filled square means that the emotion is neutral and a not filled square means that the person has a negative value from an emotion. Each color is related to one emotion: a red square is related to *Anger*, an yellow square means *Fear*, green indicates *Happiness* and blue, *Sadness*.

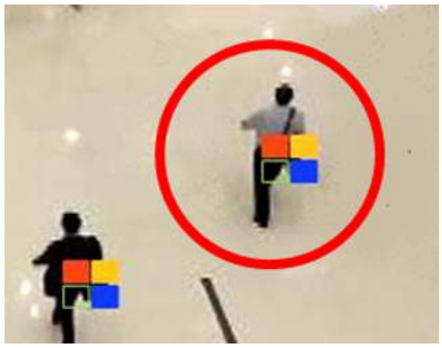
In another example, showed in Figure 5.7(a), we highlight two different situations, a group (green circle) and an individual alone (red circle). It is interesting to notice that individuals who are part of a bigger group or have a high collectivity tend to be happy, as we can see in the highlighted group in Figure 5.7(b). On the other hand, individuals who are alone and distant from others tend to experience negative emotions (see an example in Figure 5.7(c)).

It is important to keep in mind that in our approach we are not considering the internal or intrinsic emotions of pedestrians and groups, we are considering these emotions related to the physical space at a specific time.


(a) Emotion of each individual in the scene



(b) Group



(c) Individual

Figure 5.7: **Emotion analysis in the video** *CN*03 **from China**: emotion of each individual in the scene (a), focused on a group (b) and one individual (c) highlighted. Based on our approach, individuals in the group (b) tend to be happy, while the individual alone (c) tends to experience negative emotions.

### 5.3.3 Comparison with Literature

We evaluated our method in a set of 20 videos from 4 countries (9 from Brazil, 5 from China, 3 from Austria and 3 from Japan). These videos, with a duration varying between 100 and 900 frames, were collected from different public databases available on the Internet, such as (Shaikh et al., 2016; Zhou et al., 2014; Rodriguez et al., 2011). Firstly, we get the OCEAN of each individual in the scene (Figure 5.8 shows some examples). In Figure 5.8 (a) we can observe the higher E that was found in an individual, part of a group of people, while the opposite happens in (b) when lower E was computed for individual alone and far from the others.

(a) Higher E (Brazil)

(b) Lower E - Higher N (China)

(c) Higher A (Brazil)

(d) Lower A (China)

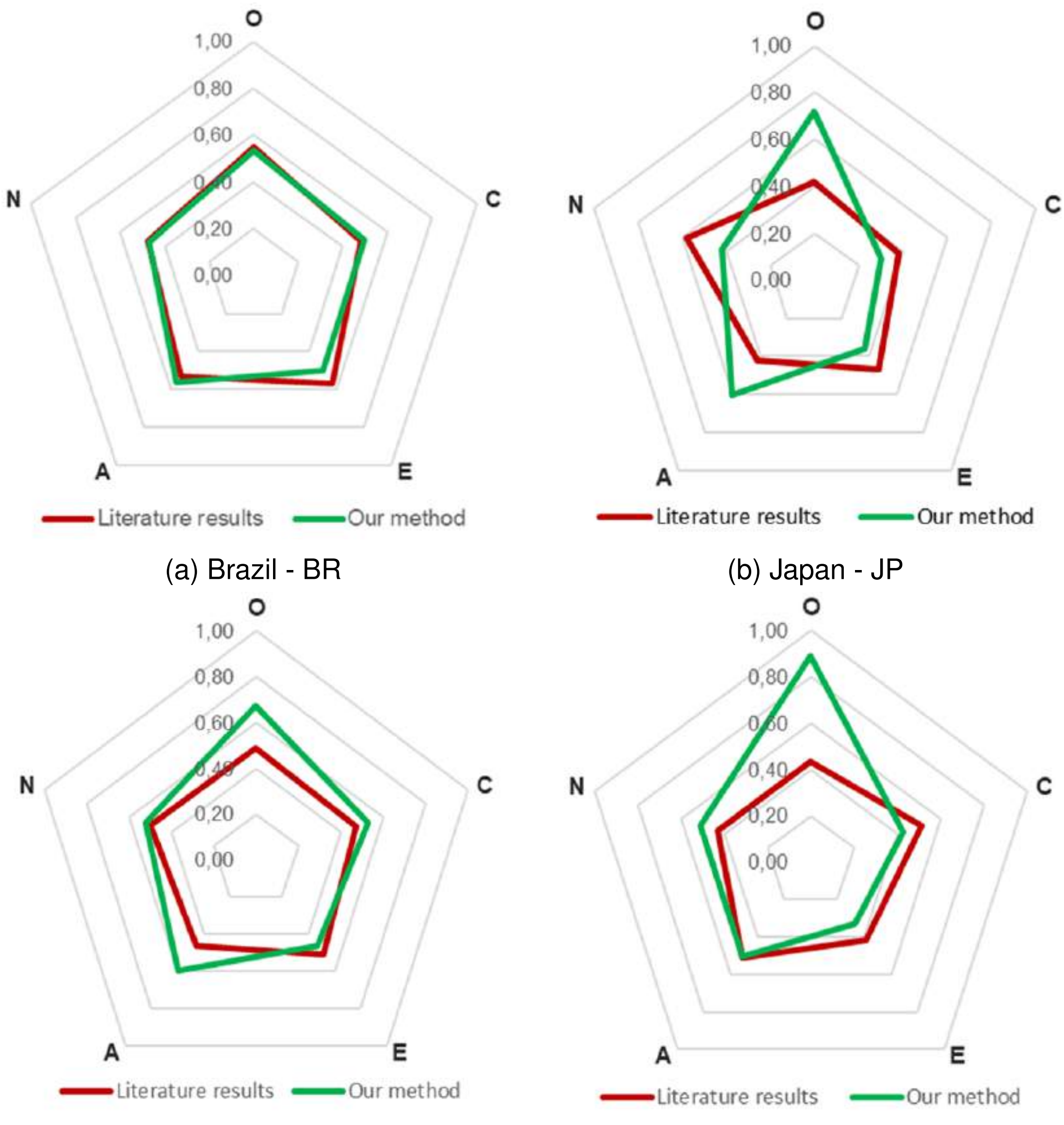
(e) Higher O (Japan)

(f) Lower O (Brazil)

Figure 5.8: **Examples of individuals OCEAN levels**: a) the highlight person has the highest Extraversion, b) shows the person with the lowest Extraversion (and highest Neuroticism), c) shows the person with the highest Agreeableness and the person highlighted in d) has the lowest Agreeableness. The highlight person in e) has the highest Openness and the person highlighted in d) has the lowest Openness.

Same kind of analysis can be done for images (c) and (d) relating to their collectivity (higher and lower respectively). Although it is more difficult to visual inspect the dimensions O, C and N we present the qualitative results. For example in Figure 5.8 (e) the highlighted individual has lower angular variation in comparison to all others (higher O value), while in (f) this is the individual with higher angular variation, consequently having lower value of O.

In addition, in Figure 5.8 (b) we obtained the higher value of N, since it is dependent of the inverse of collectivity and socialization. Once the individual OCEAN values are computed, we get the mean OCEAN value for each video. The country's OCEAN, in turn, is calculated by the average OCEANs of that country's videos. Figure 5.9 shows the results obtained by the proposed approach in all OCEAN dimensions, in comparison with the literature values presented by Costa and McCrae (2007), considered as ground-truth.




(a) Brazil - BR

(b) Japan - JP

(c) Austria - AT

(d) China - CN

Figure 5.9: **OCEAN analysis**: comparison between the proposed approach and literature values (Costa and McCrae, 2007), from Brazil (a), Japan (b), Austria (c) and China (d).

It is interesting to highlight that results achieved for Brazil (Figure 5.9(a)) showed the higher accuracy, when compared to the other countries. This was the country with more available videos to be processed in the proposed method (9 videos), in comparison with other countries.

In addition, we computed the perceptual error when accumulating each dimension from all videos and compared with literature for those Countries. Figure 5.10 shows such errors and also indicates that with an error of 18%, the dimension $E$ has the lowest value; that is an interesting observation since this was the dimension that had more questions to be analyzed, as shown in Equations 5.4, 5.5 and 5.6. The other dimensions (O, C, A and N) presented, respectively, 46%, 26%, 37% and 22%.

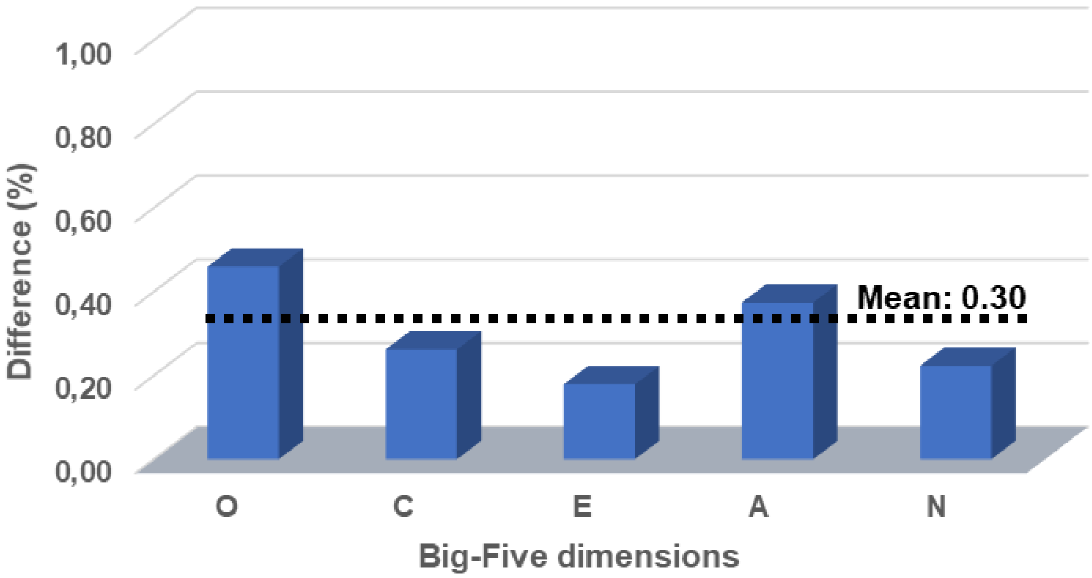



Figure 5.10: **OCEAN percentage of differences**: percentage of differences between the proposed approach and literature values, computed for all dimensions.

It is important to note that the mapping to OCEAN dimensions was empirically defined through equations using data extracted from computer vision. NEO PI-R (Costa and McCrae, 2007) measured these dimensions by considering a different type of information (subjective responses of individuals collected through questionnaires). In this sense, it is possible to affirm that, even with few videos used, the results obtained with the proposed approach are coherent with NEO PI-R results

In the next section we present an experiment which aimed to analyze the perception of some emotions and personalities by users.

## 5.4 Perception Analysis Experiment

In this section we describe an user study we performed for evaluating the perception of personality and emotion traits in crowd videos. We used a simulation environment to generate some short video sequences of pedestrian views in low density crowds. To do this, we used as a basis the tracking files of the pedestrians from the spontaneous videos and replaced the pedestrians by identical humanoids, in order to not influence the perception of the users.

In each video sequence we highlighted two individuals with different colors (red and yellow) and we asked the subjects about them. Table 5.3 shows the questions we asked to the participants with the possible answers, where the correct answer of each question is highlighted in bold.

Regarding the participants, an amount of 73 people volunteered for the experiment: 45 males (61.6%) and 28 (38.4%) females. Of these, 47.9% have some undergraduate degree. With regard to the age, 58.7% are between 20 and 30 years old, 30.7% are between 31 and 50 years old and the others 10.6% are younger or older.

Figure 5.11 shows the initial and final frames from the video $P01$, where it is possible to see a group of pedestrians in the right part of the video. Pedestrian highlighted in yellow is part of this group and the pedestrian highlighted in red walk trough the group with a higher speed. In the questions $Q1$ and $Q2$ (related to the video $P01$) we asked about which pedestrian (yellow or red) was, respectively, neurotic and angry. Figure 5.12 shows the answers given by the participants.

It was interesting to see that a little bit more than half of participants (57% in $Q1$ and 59% in Q2) answered according to the results obtained with the method proposed in this thesis. The pedestrian highlighted in red was the most neurotic and angry, according to our approach. Only a few participants answered that the pedestrian highlighted in yellow was neurotic and scared (12% in question $Q1$ and 9% in question $Q2$) and 18% answered that “neither of them” was neurotic. As we proposed in our model, geometrically, a neurotic person remains isolated and few collective. So, subjects who do think that no agent was neurotic was certainly thinking about the psychological point of view, while we are analyzing based on space relationship. In video $P01$, the pedestrian highlighted in red has these characteristics. The pedestrian highlighted in red is angry: isolated, low angular variation, low

Table 5.3: **Survey on perception of emotion and personalities**: questions and possible answers used in the research. The correct answer in highlighted in bold.

| Question | Possible answers |
|---|---|
| **Q1**: In your opinion, which of the two pedestrians highlighted in the video has a neurotic personality, yellow or red? | *a*) Yellow pedestrian;<br>*b*) **Red pedestrian**;<br>*c*) Both pedestrians;<br>*d*) Neither of them;<br>*e*) I don't know. |
| **Q2**: In your opinion, which of the two pedestrians highlighted in the video is angry, yellow or red? | *a*) Yellow pedestrian;<br>*b*) **Red pedestrian**;<br>*c*) Both pedestrians;<br>*d*) Neither of them;<br>*e*) I don't know. |
| **Q3**: In your opinion, which of the two pedestrians highlighted in the video is more openness to experiences, yellow or red? | *a*) **Yellow pedestrian**;<br>*b*) Red pedestrian;<br>*c*) Both pedestrians;<br>*d*) Neither of them;<br>*e*) I don't know. |
| **Q4**: In your perception, which of the two pedestrians highlighted in the video is afraid, yellow or red? | *a*) Yellow pedestrian;<br>*b*) **Red pedestrian**;<br>*c*) Both pedestrians;<br>*d*) Neither of them;<br>*e*) I don't know. |
| **Q5**: In your opinion, which of the two pedestrians highlighted in the video is happier, yellow or red? | *a*) **Yellow pedestrian**;<br>*b*) Red pedestrian;<br>*c*) Both pedestrians;<br>*d*) Neither of them;<br>*e*) I don't know. |
| **Q6**: In your opinion, which of the two pedestrians highlighted in the video is more extroverted, yellow or red? | *a*) **Yellow pedestrian**;<br>*b*) Red pedestrian;<br>*c*) Both pedestrians;<br>*d*) Neither of them;<br>*e*) I don't know. |
| **Q7**: In your opinion, which of the two pedestrians highlighted in the video seems to be more sociable, yellow or red? | *a*) **Yellow pedestrian**;<br>*b*) Red pedestrian;<br>*c*) Both pedestrians;<br>*d*) Neither of them;<br>*e*) I don't know. |

speed, low socialization and low collectivity. More than 50% of people chosen the red (correct) pedestrian.

Following the analysis, video $P02$ (illustrated in Figure 5.13) has a pedestrian highlighted in yellow interacting with a group of individuals and a pedestrian highlighted in red who is alone and not interacting with anyone. Questions $Q3$

(a) Initial frame (b) Final frame

Figure 5.11: **Representative frames from video** $P01$: initial (a) and final (b) frames from video $P01$ associated to questions $Q1$ and $Q2$.

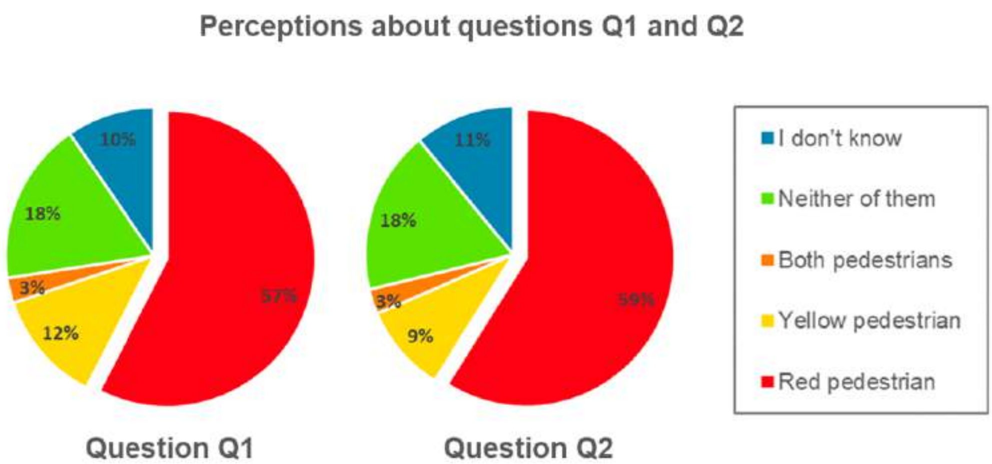
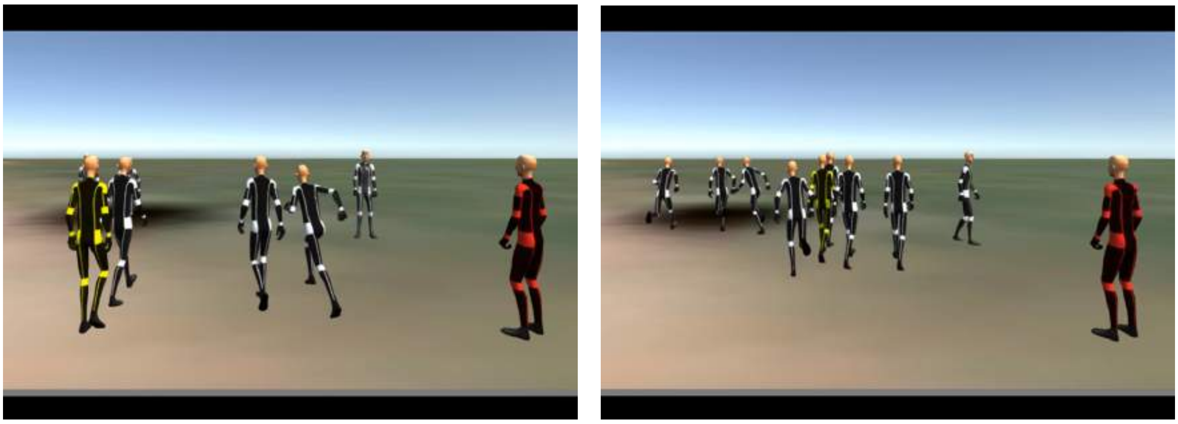



Figure 5.12: **Perception analysis concerning questions Q1 and Q2**: Question $Q1$ is related to the personality neuroticism and question $Q2$ is related to the Angry emotion.

and $Q4$, who were related to this video, asked participants about which highlighted pedestrian was, respectively, openness to experiences and afraid. Figure 5.14 shows the answers for that questions.

The results plotted in Figure 5.14 show that most of the participants perceived the same personality (in case of question $Q3$) an the same emotion (question $Q4$) that we found in our approach, i.e. 60% of the participants correctly chose the pedestrian in yellow as the most opened to experiences in question $Q3$ and 59% correctly chose the pedestrian in red as having fear. In our approach, a pedestrian



(a) Initial frame (b) Final frame

Figure 5.13: **Representative frames from video** *P*02: initial (a) and final (b) frames from video *P*02 associated to questions *Q*3 and *Q*4.

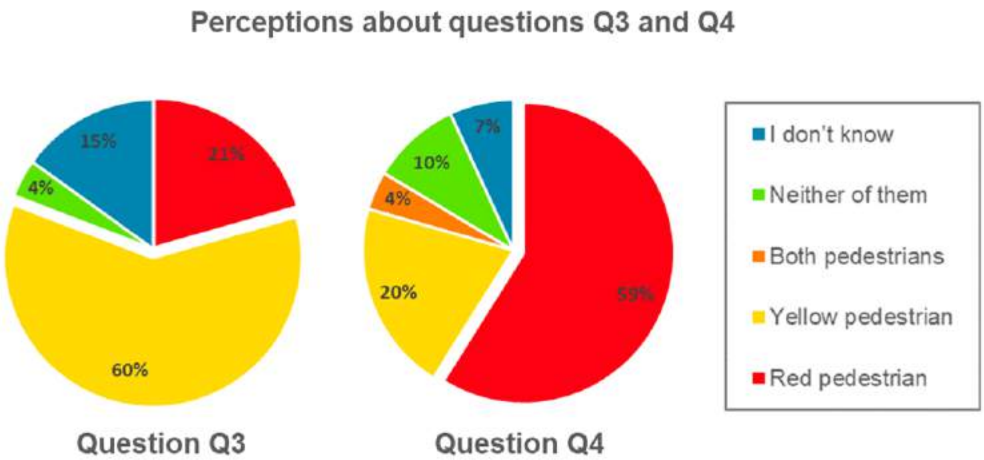
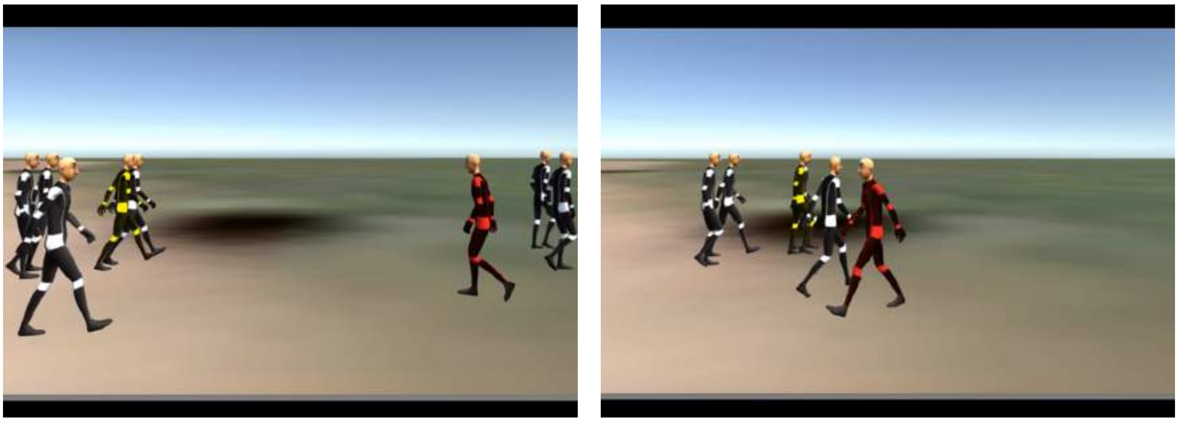



Figure 5.14: **Perception analysis concerning questions Q3 and Q4**: Question *Q*3 is related to the personality openness and question *Q*4 is related to the Fear emotion.

opened to new experiences is related to a high value for the angular variation feature. Geometrically speaking, according to what has been proposed in our model, a person who allows himself/herself to change objectives (direction) while walking is more subject to new experiences. Fear, in turn, is linked to the fact that the person is isolated from others and walks at lower speeds.

Finally, related to the video *P*03 we propose questions Q5, Q6 and Q7, asking, respectively, about happiness, extraversion and sociability. The video *P*03 (illustrated in Figure 5.15) contains a pedestrian highlighted in yellow walking with a group of people and a pedestrian highlighted in red walking alone, in the opposite direction of all other pedestrians. Regarding question *Q*5 (plotted in the left side of Figure 5.16), 40% of participants answered in accordance with the proposed approach. Geometrically, a happy person is not isolated and can present high levels of collectivity and socialization. Pedestrian highlighted in yellow presented that characteristics and was correctly identified by the participants in the survey.



(a) Initial frame

(b) Final frame

Figure 5.15: **Representative frames from video** *P*03: initial (a) and final (b) frames from video *P*03 associated to questions *Q*5, *Q*6 and *Q*7.

Questions *Q*6 and *Q*7 analyze, respectively, extraversion and sociability. In question *Q*6, although most of the participants (33% of them) correctly answer that the pedestrian highlighted in yellow is the most extrovert, it seems that the participants were not very sure about perceiving this characteristic. 25% of them answered that none of the pedestrians were extroverted, 19% replied that the most extroverted pedestrian was the one highlighted in red, 14% did not know and 9% believed that both pedestrians were extroverted.

We believe that question *Q*6 caused a greater variety of perceptions from part of the participants due to the fact that we did not explain any concept when asking the questions, nor mentioned that the perceptions would be given from the geometric point of view, considering the position of the pedestrians in the space. Many of the participants, when questioned about extroversion, may have been influenced by the movements and appearances of the humanoids rather than the geometric features. In this sense, in question *Q*7, instead of which pedestrian was more extroverted, we asked which of the pedestrians appeared to be more sociable.

When asked which pedestrian appeared to be more sociable, in question *Q*7, most of the participants (57% of them) seemed to be more convinced that the

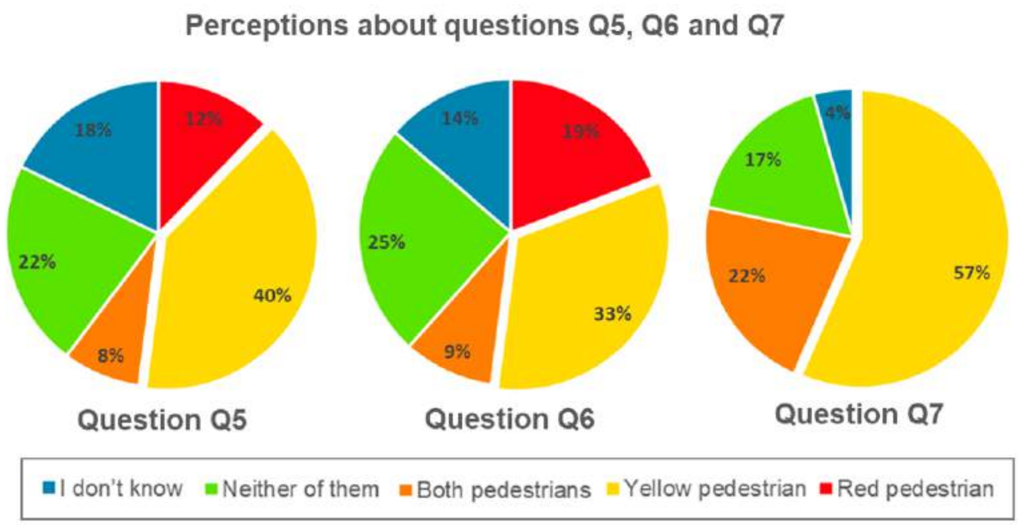



Figure 5.16: **Perception analysis concerning questions Q5, Q6 and Q7**: Question *Q*5 is related to happiness, *Q*6 is related to extraversion and question *Q*7 is related to sociability.

pedestrian highlighted in yellow is the most sociable, in accordance with the model proposed in this thesis.

## 5.5 Chapter Remarks

In this chapter we describe a way to detect individual-level traits of emotion and personality observed in video sequences, based on individuals and groups features. We propose to detect OCEAN personality traits and compared with data from different countries existent in the literature. In addition, based on OCEAN we compute four traits of emotion defined in OCC model. We believe the results are promising and video sequences can be used to detect personality and emotion, what can help us to understand people behavior in video sequences.

The results obtained from our approach indicate that our model generates coherent information when compared to data provided in available literature, as shown in various analysis. It is important to note that the mapping to OCEAN and EMOTION dimensions was empirically defined through equations using data extracted from computer vision. NEO PI-R results in the literature measured these dimensions by considering a different type of information (subjective responses of individuals collected through questionnaires). This results were published in (Favaretto et al., 2017, 2018).

These features (five dimensions of personalities and four traits of emotions) consist in our model for the third dimension *III - Personal and Emotion* geometrical dimension proposed in this thesis. Tables 5.4 and 5.5 shows a summary on how the features from previous dimensions (*I - Physical* and *II - Social*) are mapped and affect the features of, respectively, personality and emotion features.

Table 5.4: **Personality mapping influence**: influence of the features in each personality trait. (⇑) means a high value while (⇓) means a low value. (–) means that the feature does not interfere in that personality trait.

| | Ang. Variation | Collectivity | Isolation | Socialization | Speed |
|---|---|---|---|---|---|
| **O**penness (⇑) | ⇑ | – | – | – | – |
| **C**onscientiousness (⇑) | ⇓ | – | – | – | ⇑ |
| **E**xtraversion (⇑) | ⇓ | ⇑ | – | ⇑ | ⇑ |
| **A**greeableness (⇑) | – | ⇑ | – | – | – |
| **N**euroticism (⇑) | – | ⇓ | ⇑ | – | – |

For example, as described in Table 5.4, a pedestrian will get a high value in Neuroticism, if he/she gets a high value for isolation and a low value for collectivity. In a similar way, as described in Table 5.5, each personality trait interfere in any emotion. For example, a low value for Openness is related to a high value for Anger. The opposite is also true, high values for Openness imply in low values for Anger.

Table 5.5: **Emotion mapping influence**: influence of the personality traits in each emotion. (⇑) means a high value while (⇓) means a low value. (–) means that the personality trait does not interfere in that emotion.

| | **O**penness | **C**onscientiousness | **E**xtraversion | **A**greeableness | **N**euroticism |
|---|---|---|---|---|---|
| **An**ger (⇑) | ⇓ | – | ⇓ | ⇓ | ⇑ |
| **F**ear (⇑) | – | ⇓ | ⇓ | – | ⇑ |
| **H**appiness (⇑) | – | – | ⇑ | – | ⇓ |
| **S**adness (⇑) | – | – | ⇓ | – | ⇑ |

In addition, we performed an experiment to evaluate if people can perceive different personalities and emotions performed by pedestrians in crowds. We used a simulation framework based on real video tracking files and replaced the pedestrians by identical avatars, generating short video sequences that were evaluated by the participants in the survey. It was interesting to see that, even without explaining to the participants the concepts of each personality or emotion and how they were calculated in our approach (considering the geometric characteristics), in all the cases, most of the participants perceived the personality and emotion that the agent was expressing in the video, in accordance with our approach.

Next chapter is responsible for present an approach to detect cultural aspects from videos considering the Hofstede Cultural Dimensions (Hofstede et al., 1991), theses features are related to the last and fourth dimension proposed in this thesis: *IV - Cultural* geometrical dimension.

# 6. FOURTH DIMENSION: DETECTION OF CULTURAL ASPECTS IN CROWDS

In this chapter is presented the approach to detect cultural aspects from videos considering Hofstede Cultural Dimensions (Hofstede et al., 1991). In our model, the Hofstede cultural dimensions (presented in Section 2.1.1) is related to the *IV - Cultural* geometrical dimension.

In order to successfully map the characteristics of pedestrians and groups in cultural aspects, we considered some information extracted from pedestrians in the video presented in Chapter 4 (First and Second Dimensions: Data Extraction, Crowd Types and Video Similarity). For each video $v$ processed, we calculate the average of all groups present in the video, making it possible to find the final values for each dimension in percentages to compare with the Hofstede values, which are also described in percentages.

Section 6.1 presents the proposed approach to the mapping of the characteristics from the pedestrians in the video to the Hofstede Cultural Dimensions. In this sense, we use the information about groups (described in Section 4.2), detected through the pedestrian tracking. Following, Section 6.2 presents and discusses the results obtained with this approach, also comparing the values obtained with the Hofstede dimensions values.

## 6.1 The Proposed Approach

For each information extracted from the videos to be analyzed a percentage value is computed (function *fp*, a simple function to find the percentage, given the maximum value and the value obtained in each video $v$), in relation to its maximum possible values (with the exception of the average distance $\bar{d}$ between the individuals of the group). Therefore, the following information is considered:

- Mean distances between the individuals in each group $g$ averaged by all groups in video $v$, i.e. $\bar{d}_v$. Once we do not know the maximum possible value we decided to attribute a percentage based on Halls distances Hall (1966), as follows:

  we attribute $PDI_v$ = 20% if $\bar{d}_v$ corresponds to intimate distances,

$PDI_v$ = 40% if personal distances,

$PDI_v$ = 60% for social distances and

$PDI_v$ = 80% for public distances.

- Speed percentage of video $v$: $S_v = fp(\bar{s}_v, 1.4)$, where $\bar{s}_v$ states for the average of all groups $g$ speeds.
- Percentage of angular variation: $O_v = fp(\bar{\alpha}_v, 360°)$, where $\bar{\alpha}_v$ states for the average of all groups $g$ $\alpha_g$; and
- Percentage of group cohesion: $GC_v = fp(\bar{C}_v, 3)$, where $\bar{C}_v$ states for the average of all groups $g$ cohesion values.

All these values were empirically defined, in particular, the maximum cohesion value was based on the approach proposed by Bassi (2006).

### 6.1.1 Mapping Group Characteristics to Cultural Dimensions

Before proposing a metric for each HCD dimension, we looked at research in different areas of knowledge. The dimension *IDV* (individualism vs. collectivism) was considered in the work proposed by Lala et al. (2011) as a function of the interpersonal space and the speed of individuals. In this work, this dimension is calculated directly based on the quantities of people alone and grouped in the video. Thus, collectivism ($COL_v$) of video $v$ is the percentage of people grouped while individualism ($IDV_v$) is the percentage of people alone.

The dimensions *PDI* (Power distance index), *MAS* (Masculinity vs. Femininity), *LTO* (Long vs. short term orientation) and *IND* (Indulgence vs. Moderation) dimensions were not considered in previously work, as far as we know. Thus, in this work an empirical parameterization is proposed to consider these characteristics, as following described.

In relation to the *PDI* dimension, which represents the differences of status between individuals and the behavior of these groups in society, the hypothesis is that individuals who stay close to each other recognize less the group hierarchy, whereas large distances among the individuals may represent a more explicit recognition of the hierarchy. In this way, each group average distance ($\bar{d}_g$) is computed to determine the $(\bar{d}_v)$ of video $v$, and this value is used together with Hall distances Hall (1966) to find $PDI_v$ for video $v$.

In terms of $LTO_v$ of video $v$, where the idea is persistence (long term) as opposed to fast results (short term), in this work the group orientation $O_v$, as explained before, was adapted to be mapped to this dimension. Thus, high values of angular variation result in short-term orientation ($STO = 100 - LTO$, where $LTO$ is calculated according to Equation 6.1).

$$LTO_v = \begin{cases} O_v, & \text{if } O_v >= 50 \\ 100 - O_v, & \text{otherwise.} \end{cases} \tag{6.1}$$

Considering the $MAS_v$ dimension, the “preference for cooperation” we considered that this dimension can be defined in terms of average groups cohesion found in the video $v$. High levels of cohesion represent greater femininity in this dimension, while low values of cohesion indicate greater masculinity. The value of $LTO_v$ was used as a weight to regulate the value of $MAS_v$, as defined in Equation 6.2:

$$MAS_v = \chi_1 GC_v + (1 - \chi_1) LTO_v, \tag{6.2}$$

where $\chi_1 = 0.5$ is an empirically defined constant.

Finally, the $IND$ dimension was characterized by the average velocity and collectivity of the groups in video $v$, determined by the Equation 6.3:

$$IND_v = \rho_1 S_v + (1 - \rho_1) COL_v, \tag{6.3}$$

where $\rho_1 = 0.5$ is also an empirically defined constant.

It is important to note that, in our proposed mapping of cultural dimensions, some equations have been defined to quantify the cultural dimensions of Hofstede (Hofstede, 2001) through computer vision. On the other hand, Geert Hofstede measured these dimensions by considering a different type of information (subjective responses of individuals collected through questionnaires). Our hypothesis is that there is a relationship about how people answer questions and how they behave in the space/time as mentioned earlier. In the next section are presented some results obtained in that analysis.

## 6.2 Results obtained

In this section we discuss the results obtained with the proposal of culturality mapping presented in the previous section. Initially, an experiment was carried out with three countries, involving 14 videos (seven videos from Brazil, four videos from China and three videos from Austria).

The first step of the experiment was to analyze the detection of groups. The reference value (ground truth, manually annotated) corresponding to the total of groups present in these videos is 83. The proposed model correctly detected 71 of them (85%). From those detected incorrectly (15%), two were false-positive and ten false-negative.

Based on the behavior of these groups and their parameters, the mapping of culturality was carried out and compared with the results of the Cultural Dimensions of Hofstede (HCD). Figure 6.1 shows the values of the cultural dimensions computed with the proposed method and the values present in HCD[1] (Hofstede, 2011).

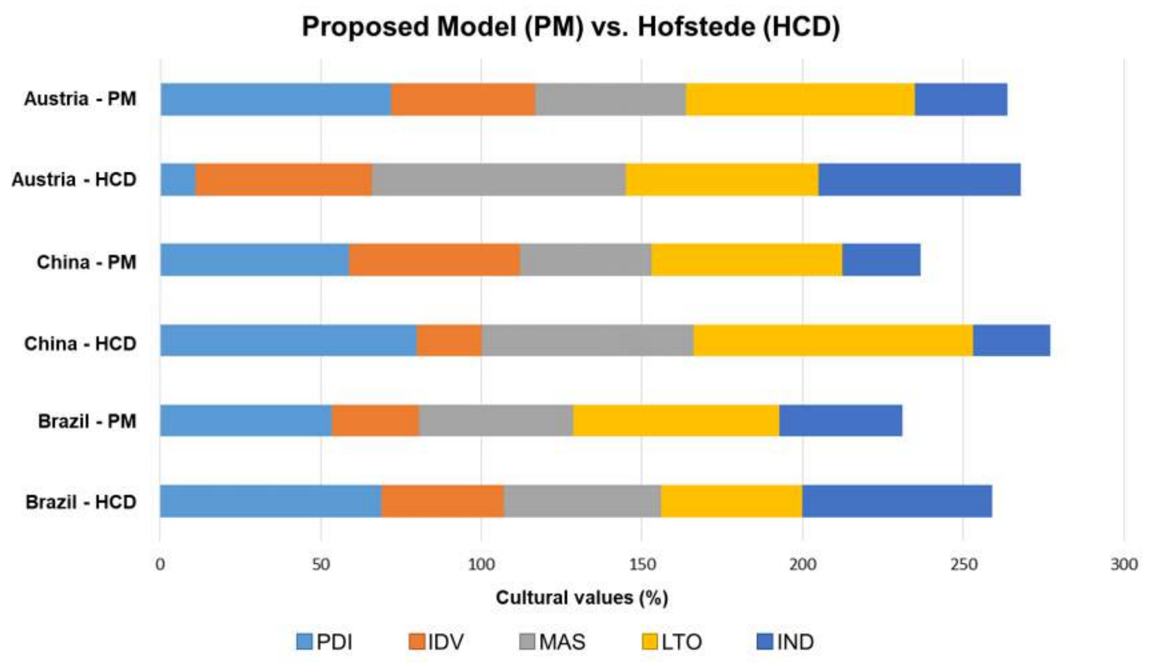



Figure 6.1: **Our approach vs. Hofstede**: comparison of the values (%) obtained in each cultural dimension.

[1]The HCD values from many countries can be find at: http://geert-hofstede.com/

It is important to note that we can not use HCD as ground truth for the proposed method, since they deal with different information (subject answers in HCD and images in the proposed method); however, it is possible to evaluate if results are coherent to be used to detect cultural aspects in different countries video sequences. The obtained average differences for each country presented in Figure 6.1, considering all the cultural dimensions, are 13.70%, $21,53\%$ and 29.77% for Brazil, China and Austria, respectively. So, we consider that even if we used only a few videos, our results are coherent with HCD.

In order to verify if the proposed model allows to detect cultural differences from videos of crowds in order to classify them, the average distances of the five dimensions were computed for each of the seven videos of Brazil and compared to the averages of the three countries: Brazil, China and Austria. For instance, in Figure 6.2, video 1 from Brazil is less different from all Brazilian videos that from other countries.

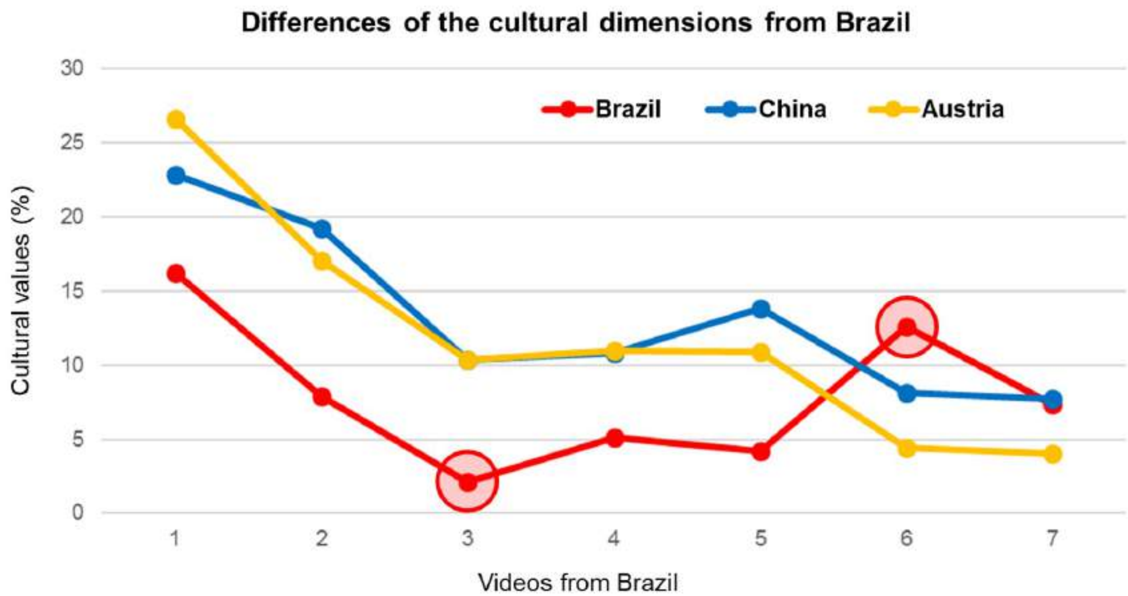



Figure 6.2: **Differences from Brazil**: differences from seven videos made in Brazil and the three main groups: Brazil, China and Austria.

The last two videos from Brazil ($v = 6$ and $v = 7$) could be erroneously classified as China or Austria. In fact, these two videos presented higher values in both dimensions: IDP (19.84% and 17.55%) and IDV (30.50% and 20.50%), which were higher than the average of Brazilian videos. The third Brazilian video ($v = 3$, highlighted in Figure 6.2) obtained the smallest difference in relation to the average of Brazilian videos.

The video $v$ = 3 is illustrated in Figure 6.3(a), where it is possible to see the high number of groups[2]. On the other hand, the video $v$ = 6 (also highlighted in Figure 6.2 and represented in Figure 6.3(b)), illustrates more people alone and only one permanent group by the blue circle, appears only at the end of the video, not having time to stabilize and be detected as permanent).

(a) Brazil $v$ = 3 (b) Brazil $v$ = 6

Figure 6.3: **Scenes from the 3rd and 6th videos from Brazil**: largest number of groups in the video $v$ = 3 (a) compared to the video $v$ = 6 (b).

In a last analysis involving these videos, the three Countries (Brazil with 7, China, 4 and Austria with 3 videos) are culturally compared. Figure 6.4 illustrates the average of each cultural dimension of the three countries of the experiment, with a highlight for the lowest *IDV* (Brazil), as the opposite of China, and the largest *LTO* (Austria).

Figure 6.5 illustrates the results of the cultural dimensions obtained with the proposed method in comparison to the Hofstede values. The experiment was carried out with 14 countries present in the *Cultural Crowds* dataset[3].

In Figure 6.5(b), the values of the dimensions *IND* and *LTO* on the Hofstede website (https://geert-hofstede.com) were not available for the country Arab Emirates, which is why the results of these two dimensions were omitted in the proposed method (*IND* = 0 and *LTO* = 0). The goal of these plots is not to use the Hofstede values as ground truth, but to verify if the results are coherent with each other, since they come from different sources (self-report in Hofstede and computational vision in the proposed method). France (FR) and Portugal (PT) presented the

[2]The people in the upper right corner of the video are walk closing to the others, the group does not stabilize, making it impossible to be detected by the proposed model.

[3]The *Cultural Crowds* dataset is described in Appendix A of this manuscript

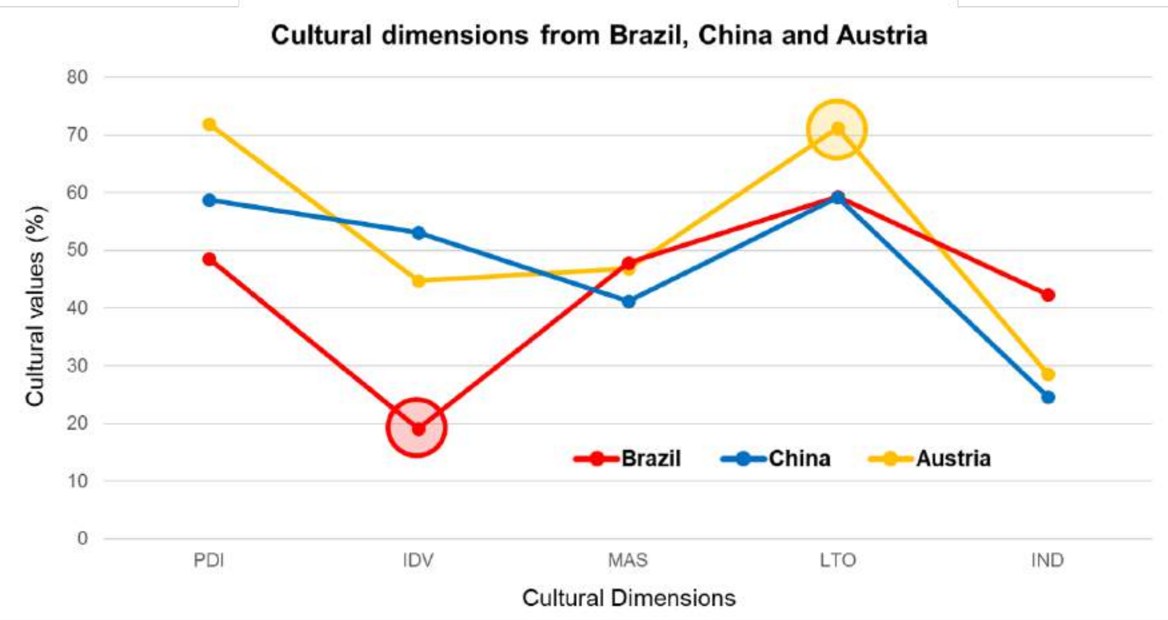
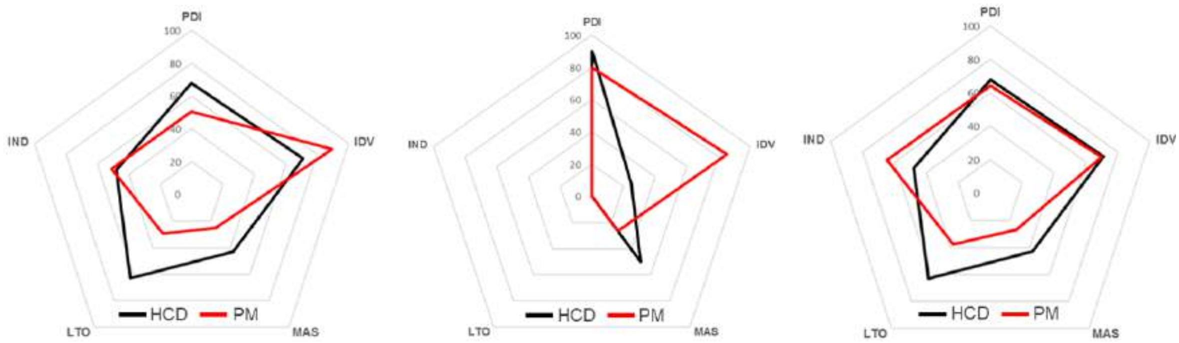
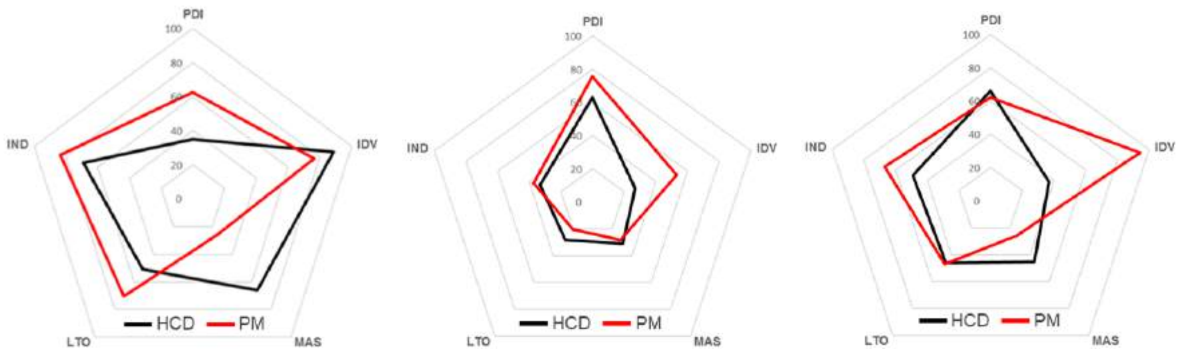



Figure 6.4: **Cultural comparison among the three countries analyzed**: Brazil, China and Austria. We highlight the *IDV* which is lower in Brazil, and *LTO* that is higher in Austria




(a) Germany - *DE* (b) Arab Emirates - *AE* (c) France - *FR*




(d) England - *UK* (e) Portugal - *PT* (f) Turkey - *TR*

Figure 6.5: **Proposed method vs. Hofstede**: results of culturality for each of the experiment countries. Hofstede (HCD, black line) versus the proposed method (PM, red line).

smallest differences in relation to the values obtained by Hofstede. The values of these countries are shown, respectively, in Figure 6.5(c) and (d).

In order to verify if the results obtained with the proposed method are consistent with Hofstede, the average difference of culturality for each of the dimensions was computed considering all countries involved in the experiment. In this analysis, illustrated in Figure 6.6, the results of all videos were considered in the calculation of the mean of each dimension.

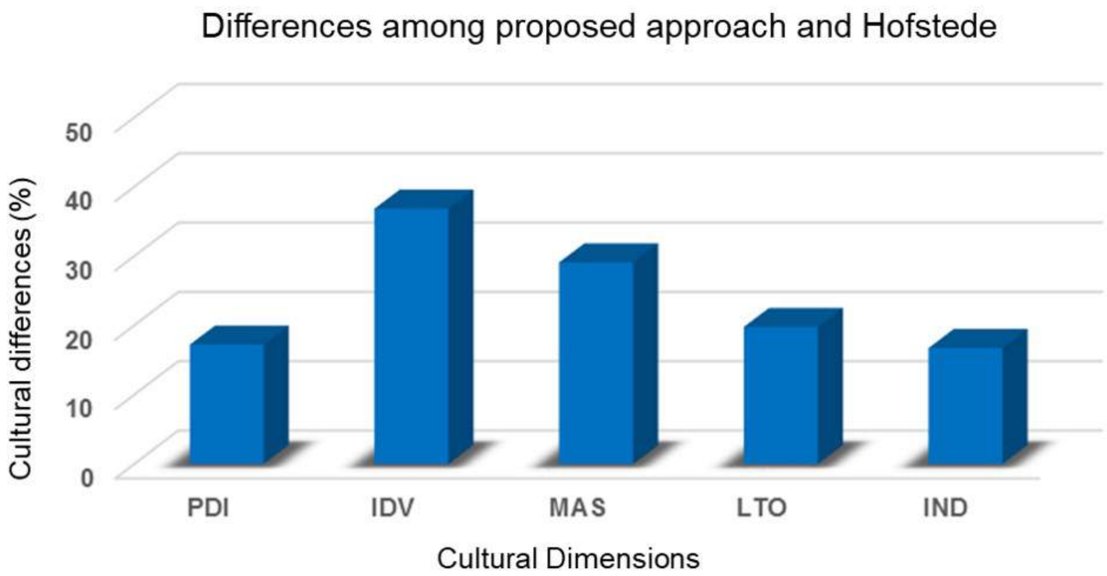



Figure 6.6: **Proposed method versus Hofstede**: mean of culture differences for all cultural dimensions, considering all countries involved in the experiment.

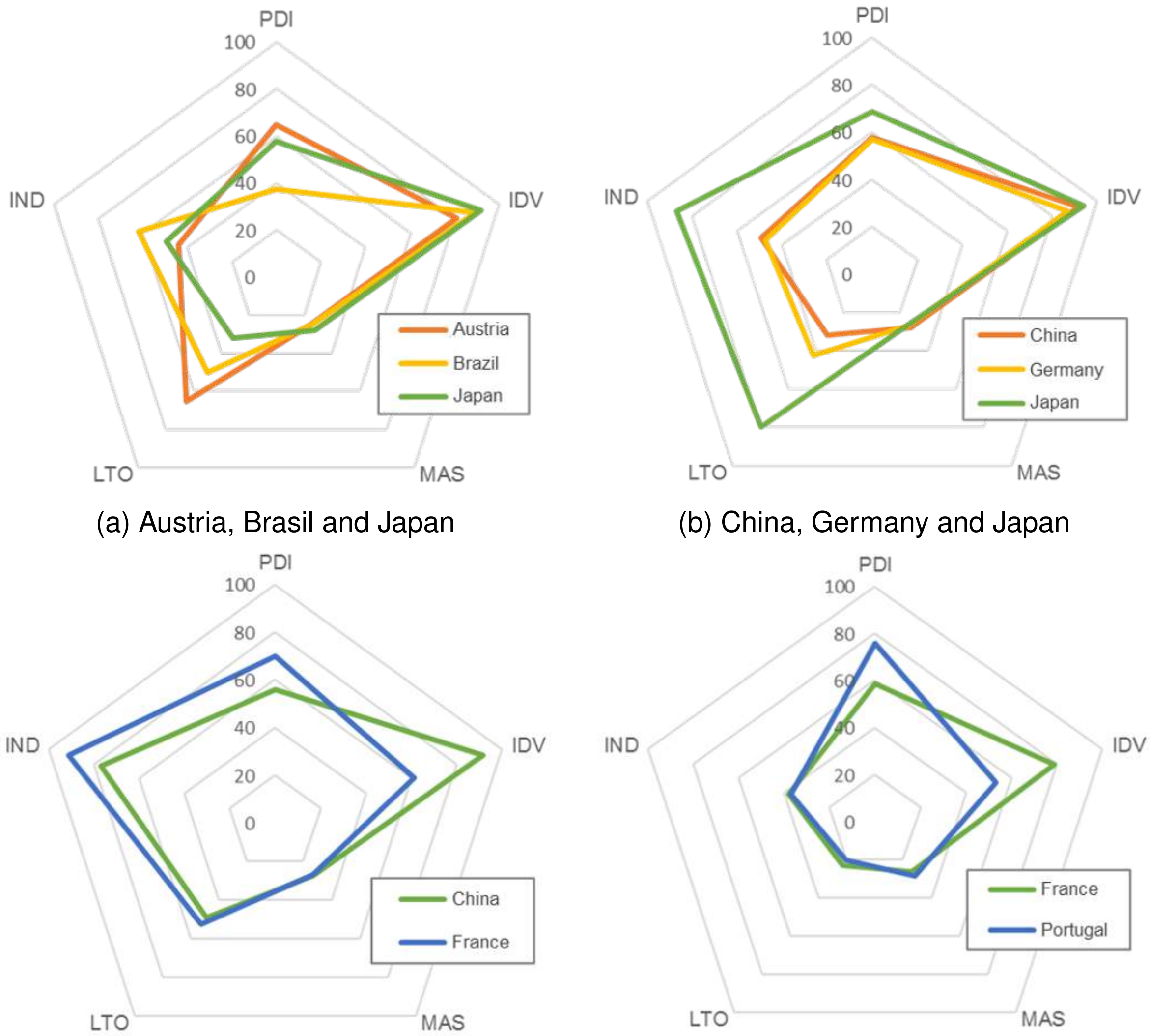
With an average of 36.83%, the *IDV* dimension (Individualism vs. collectivism) was the one that most distanced dimension if compared to Hosftede values. Analyzing Figure 6.5, although in some countries like France (Figure 6.5(c)) and England (Figure 6.5(d)) the value for this dimension is very close to that obtained by Hofstede, in others it is very distant, contributing to the increase of the difference, as is the case of Arab Emirates (Figure 6.5(b)) and Turkey (Figure 6.5(f)).

The *PDI* dimension, on the other hand, with 16.72% of difference, was the dimension who presented the lowest distance to Hofstede's results, suggesting that the use of the proximity of individuals to infer the level of hierarchy of a group may have been a good hypothesis.

Figure 6.7 shows an analysis of culturality considering similar crowds, but from different countries. This similarity was obtained according to the method proposed and presented in Chapter 4, consisting of the following videos:

- Austria (AT-01), Brazil (BR-02) and Japan (JP-02), illustrated in Figure 6.7(a);
- China (CN-05), Germany (DE-02) and Japan (JP-03), illustrated in Figure 6.7(b);
- China (CN-04) and France (FR-02), illustrated in Figure 6.7(c); and
- France (FR-01) and Portugal (PT-01), illustrated in Figure 6.7(d).

The frames illustrating the crowds of each of the videos mentioned in Figure 6.7 can be seen in the Section *Cultural Crowds Dataset* (showed in Figure A.2), referenced by the indicators in parentheses.




(a) Austria, Brasil and Japan

(b) China, Germany and Japan

(c) China and France

(d) France and Portugal

Figure 6.7: **Cultural aspects in similar crowds**: (a) Austria (AT-01), Brazil (BR-02) and Japan (JP-02), (b) China (CN-05), Germany (DE-02) and Japan (JP-03), (c) China (CN-04) and France (FR-02) and (d) France (FR-01) and Portugal (PT-01).

It is possible to notice in Figure 6.7 that, although they are similar crowds (number of people and groups, sizes of groups, among others), there are differences in some dimensions, which vary according to the country. In Figure 6.7(b), for example, the Japanese crowd has much higher values for the *PDI*, *IND* and *LTO* dimensions, while the crowds from China and Germany are quite similar.

In the comparison among France and Portugal crowds, represented in Figure 6.7(d), the two dimensions that differ are *PDI* and *IDV*. In relation to *PDI*, Portugal presented a higher index because the average distance between the people of the groups was greater than in France. In the case of the *IDV*, France had a higher rate due to the greater number of people alone in its videos, compared to the video of Portugal.

Brazil stood out in the *PDI* dimension when compared to the crowds of Austria and Japan, as illustrated in Figure 6.7(a). In this case, the lowest value for this dimension is related to the fact that Brazil has larger groups, with people walking closer than in the other two countries.

With respect to the comparison between the crowds of China and France, represented in Figure 6.7(c), the highlight of China's highest value in the *IDV* dimension is due to the fact that this country's crowd has a larger number of people not grouped in comparison to the French crowd.

## 6.3 Big4GD Dimensions

With the content of this chapter we conclude the presentation of the **Big-Four Geometric Dimensions**, or just **Big4GD** model, illustrated in Figure 6.8. This four dimensions of features aims to characterize pedestrians organized in groups and crowds with regard to space and time relation, i.e. from the geometrical point of view.

In summary, the *I – Physical* deals with physical features of pedestrians obtained directly from the tracking, such as speeds, distances from a pedestrian from others, angular variations, among others; *II – Social* dimension is related to the social interactions, characterizing groups of pedestrians and social features, as collectivity, isolation and socialization levels of individuals; *III – Personal and Emotional* dimension maintains the features related to personality (Big-Five) and emo-

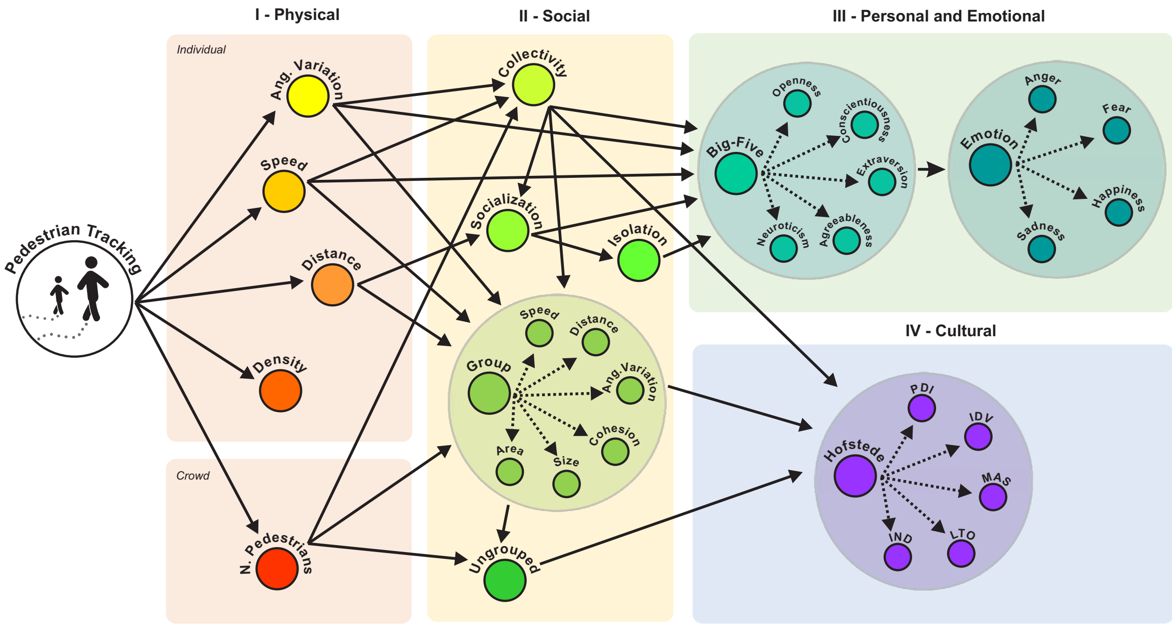



Figure 6.8: **Our Big4GD model**: features mapping from the proposed approach divided into four dimensions: *I – Physical*, *II – Social*, *III – Personal and Emotional* and finally, *IV – Cultural*.

tion (OCC) traits; and *IV – Cultural* dimension deals with features regarding cultural aspects, according to Hofstede (HCD).

In previous chapters we present how each of the pedestrians and group features were defined and how these features are organized in the proposed four-dimensional model. In addition, a set of experiments was presented with the purpose of analyzing and validating the proposed model.

## 6.4 Chapter Remarks

In this chapter we presented how we proceed to characterize groups of pedestrians in video sequences and detect some cultural aspects. We were inspired in cultural dimensions proposed by Hofstede. Even with the differences regarding the used method to capture information (subject answers and computer vision), the results indicate a coherence in our mostly empirical equations. This results were published in (Favaretto et al., 2016a).

Table 6.1 shows the influence of the features in each Hofstede dimension. (⇑) means a high value, (⇓) means a low value and (–) means that the feature does not interfere in that dimension. For example, high values for mean distance among the pedestrians from a group cause a high value in the *PDI* dimension.

Table 6.1: **Hofstede mapping influence**: influence of the features in each Hofstede dimension. (⇑) means a high value while (⇓) means a low value. (–) means that the feature does not interfere in that dimensions.

| | M. distance | M. ang. var. | M. speed | Cohesion | Ung. pedestrians |
|---|---|---|---|---|---|
| *PDI* (⇑) | ⇑ | – | – | – | – |
| *IDV* (⇑) | – | – | – | – | ⇑ |
| *MAS* (⇑) | – | ⇑ | – | ⇓ | – |
| *LTO* (⇑) | – | ⇑ | – | – | – |
| *IND* (⇑) | – | – | ⇓ | – | ⇑ |

The results obtained in the experiments performed in this chapter indicate that the proposed method seems to be promising when used to classify videos from Brazil, based on the similarity of the cultural dimensions, as shown in Figure 6.2.

In addition to the results already discussed in last chapters, in the next chapter we present some analysis in a controlled experiment. For that, we use Fundamental Diagrams to compare pedestrians from three countries: India, Brazil and Germany. Due to the experiment characteristics, we evaluate our model with experiments involving only individuals and not groups. In this sense, the *IV – Cultural* dimension was not evaluated in that experiment. In addition, we use simulation to reproduce the patterns observed in FD experiment. We also performed a comparison with the work proposed by Sorokowska et al. (2017), concerning distances between pedestrians.

# 7. FUNDAMENTAL DIAGRAM ANALYSIS

In this chapter, we evaluated our model considering the Fundamental Diagrams experiment regarding personal distance and walking speeds in three different countries: India, Brazil and Germany in a controlled experiment. In this evaluation, due to the performed experiment, i,e. only one group at each time executing the same task, we analyzed only individuals and not more than one group in each video, reason why the *IV – Cultural* dimension (HCD) was not evaluated. In addition, we investigate the personal distance obtained in the experiments and compare with studies performed by Sorokowska et al. (2017). As another topic, but related, we shown some complementary investigations regarding the observed FD patterns obtained in crowd simulations.

Section 7.1 presents the proposed approach we performed in the experiment. Section 7.2 is responsible for analyze and discuss some results from the FD experiment. Section 7.3 addresses an analysis involving personal space in a comparison with Sorokowska et al. (2017). Section 7.4 presents an analysis regarding personality and emotion traits in the Fundamental Diagram. Finally, Section 7.5 describe how we use two different simulation environments well known in crowd simulation area: BioCrowds and ORCA to reproduce patterns of FD in crowds.

## 7.1 Proposed approach for analysing FD

We propose a methodology based on three main modules: trajectory detection, data extraction and FD analysis. For the first component we obtain the individual trajectories of observed pedestrians from real videos using Computer Vision where a Fundamental diagram experiment is performed as illustrated in Figure 7.1.

This experiment was applied as described in Chattaraj et al. (2009) to three countries (India, Germany and Brazil) with the same populations (N=15, 20, 25, 30 and 34). A corridor was built with markers and tape placed on the ground. Its size and shape is presented in Figure 7.1. At the base of the corridor was a rectangle designating the Region of Interest (ROI) from which the populations were captured as proposed by Chattaraj et al. (2009).

The experiments conducted in the three countries differed in terms of the moving directions and compositions of the test subjects. For the German study, the

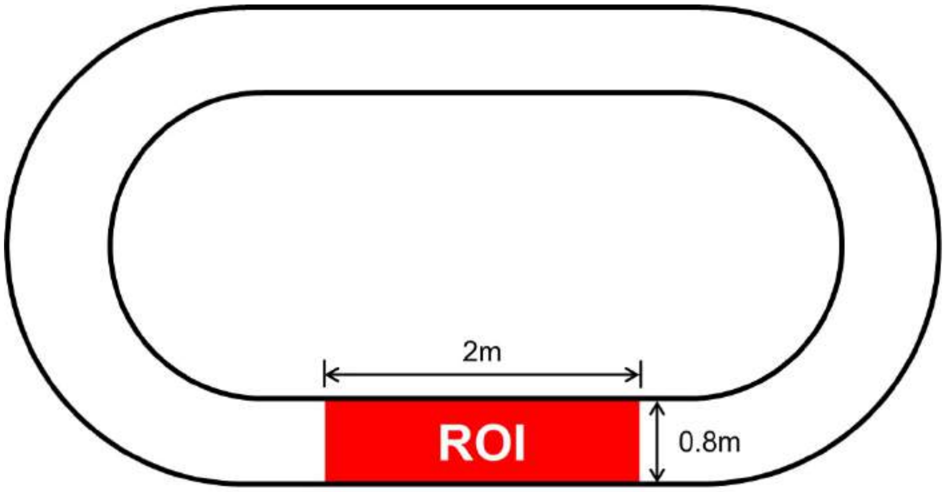



Figure 7.1: **Experimental setup**: sketch of the experimental setup used as proposed by Chattaraj et al. (2009).

group of test persons included male and female students. For the Indian study, the group consisted of only male participants. Participants of the Brazilian study were similar to those of the German study and included male and female students.

For moving directions, experiments conducted in India and Brazil followed a clockwise direction while those conducted in Germany followed an anti-clockwise directions. The experiments conducted in Germany and Brazil were performed indoors, while those conducted in India were performed outdoors on paved ground.

For the experiment performed in Brazil, we created videos of the same population occupying the same environmental setup as that used for the other countries Chattaraj et al. (2009) employing different camera positioning. The camera was positioned overtop of the participants to eliminate the video perspective while for the German and Indian experiments the camera was positioned on the ground in front of the ROI.

We used this approach to obtain pedestrian information through computational tracking while in other studies information was collected by a person as pedestrians crossed a demarcated position on the video screen. Figure 7.2 shows the experiment performed in the three countries with $N$ = 30 (where $N$ is the number of pedestrians).

For the Brazilian experiment, as was the case for the other experiments, all individuals were initially uniformly distributed throughout the corridor. After instruction was delivered, every individual was instructed to walk around the corridor twice and to then leave the environment while continuing to walk a reasonable distance,


Figure 7.2: **FD experiment performed with** $N$ = 30 **pedestrians**: **on the left**: India (Chattaraj et al., 2009), **in the centre**: Germany (Chattaraj et al., 2009) and **on the right**: Brazil.

eliminating the tailback effect. For the second component of our method, geometric information was obtained from trajectories and was analyzed to find neighbouring individuals and to compute distances among them. The third module involved fundamental diagram analysis.

### 7.1.1 Methods

Our research was carried out in accordance with relevant international and Brazilian guidelines and regulations. Research protocols used were approved of by the Scientific Review Board of the Pontifical Catholic University of Rio Grande do Sul Graduate Program in Computer Science.

Informed consent was obtained to use images that could lead to the identification of a study participant and to publish corresponding information/images in an online open-access publication. The survey was conducted in line with the research ethics committee requirements. All individuals gave written informed consent to participate in this study and they can not be recognized.

#### Trajectory Recovery

In the case of India, we do not have all videos (or tracking) generated from the experiments performed in these countries. Upon request, the authors Chattaraj et al. (2009) sent us part of the dataset, i.e., the time (in seconds) at which each individual i enters ($t_i^{in}$) and exits ($t_i^{out}$) the experiment ROI (Region of Interest). Based on this information and on the length of the ROI (2$m$ in length), we compute the speed of individual $i$ as:

$$s_i = \frac{2}{t_i^{out} - t_i^{in}}, \tag{7.1}$$

and we compute the distance between individual $i$ and his/her predecessor $i-1$ as:

$$d_{i,i-1} = s_i |t_i^{in} - t_{i-1}^{in}|, \tag{7.2}$$

where $t_i^{in}$ is the time at which individual $i$ enters the ROI and where $t_{i-1}^{in}$ is the time at which his/her predecessor $i-1$ enters.

For the Brazilian and Germany experiment, on the other hand, we use the same information available from India for each individual $i$ (i.e. $t_i^{in}$ and $t_i^{out}$) using a tracking method. Initial individual detection is performed using the method proposed by Viola and Jones (2001). The boosted classifier based on haar-like features was trained with 4500 views of individuals' heads as positive examples and with 1000 other views used as negative examples (the CoffeBreak and Caviar Head datasets from Tosato et al. (2013) were used). This detector performs initial position detection based on the positioning of individuals' heads, which are used as input parameters for the next step: tracking.

Once individuals are detected, trajectories are obtained using a method proposed by Bins et al. (2013). This approach to object tracking is based on multiple disjoint patches obtained from the target. The patches are represented parametrically by the mean vector and covariance matrix computed from a set of feature vectors that represent each pixel of the target. Each patch is tracked independently from the Bhattacharyya distance Fukunaga (1990), and the displacement of the whole template is obtained using a Weighted Vector Median Filter (WVMF). To smooth the trajectory and to address short-term total occlusions, a predicted displacement vector based on the motion of the target in previous frames is used. Appearance changes of the target are managed by an updating scheme.

As tracker input parameters, we use the initial position of heads detected in the previous step. To obtain the desired parameters for the world coordinate system, we compute the planar homography of each video and transform the extracted trajectories to the world coordinate system by assuming that the head position occupies the ground space ($z = 0$). As our videos are made from a bird's eye view, this assumption does not produce significant errors in projections. The tracker outputs individual trajectories in world coordinates (in Figure 7.3 we present a result generated from our work). Such information is used in the next step as discussed in the Section *Fundamental Diagram Data Extraction*.

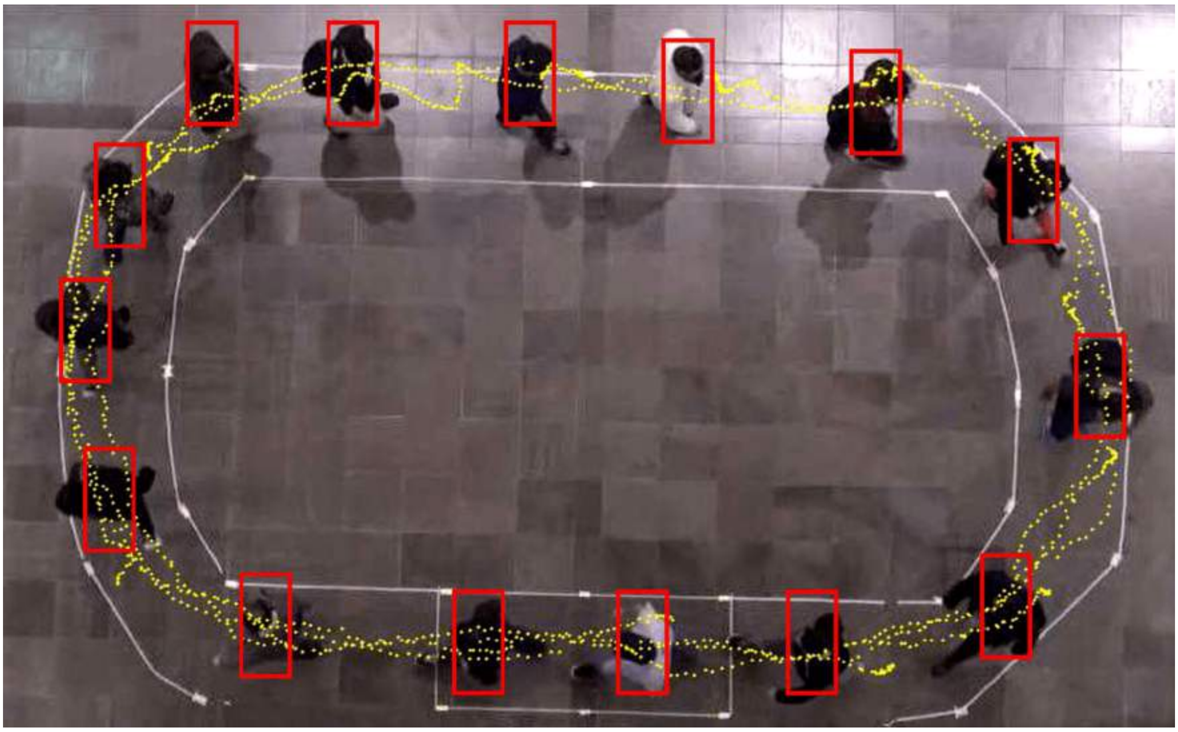


Figure 7.3: **Snapshot of tracking used for the Brazilian experiment**: trajectories (yellow dots) were obtained with the approach developed by Bins et al. (2013).

Fundamental Diagram Data Extraction

As noted above, we observed information from the authors on FDs for Germany and India. For Brazil, we applied a tracking process to extract information as described above. We computed basic information of the individual level for each person *i* and for each frame: *i)* the 2D position $\vec{X}_i$ of person *i* (meters) and *ii)* the speed $s_i$ of person *i* (meters/frame). Such data on pedestrians was used to compute the Fundamental Diagram from the rectangle of analysis. For each frame we computed the number of individuals within the ROI (2 m-long rectangle of analysis) and the mean population measured over one second (with a video acquisition frame of 30 FPS (frames per second)).

Using the same method we computed the mean density and speed of individuals from the ROI per second. In turn, we determined the density (individuals per sqm) and speed (meters per second) as described by Chattaraj et al. (2009) and as illustrated in Figure 7.4 (on the right). In this figure we plot data captured from video sequences containing 15, 20, 25, 30 and 34 individuals while walking in the experimental scenario. As we show, the Brazilian subjects travel at higher speeds than the other two samples and even at high densities (e.g., of $N = 30$ and 34).

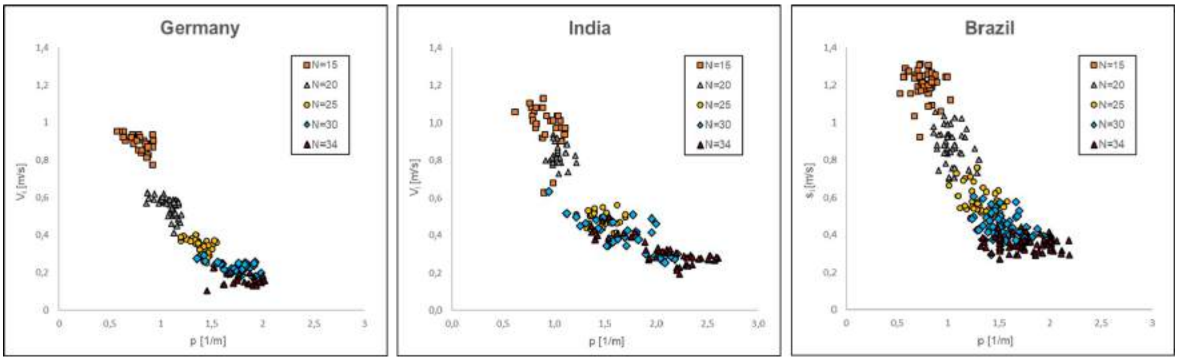



Figure 7.4: **Speed density data extracted from Fundamental Diagrams for all the three countries**: (**left**) Germany (Chattaraj et al., 2009), (**centre**) India (Chattaraj et al., 2009) and (**right**) Brazil.

As was computed for the other countries, the speed for the Brazilian experiment was calculated based on the amount of time it took for a certain pedestrian $i$ to cross a 2-m ROI, i.e., as described in Equation 7.1. The mean speed measured at each frame $t$ of the video sequence was computed as follows:

$$\bar{s}_t = \frac{1}{n_t} \sum_{p=1}^{n_t} s_{p,t}, \tag{7.3}$$

where $\bar{s}_t$ is the mean speed measured at time $t$. $n_t$ is the number of individuals examined in the rectangle analysis in $t$, and $s_{p,t}$ is the speed of person $p$ at time $t$. In a similar way, the density is computed from a 2-m-rectangle analysis as shown in Equation 7.4:

$$\bar{d}_t = \frac{n_t}{2}. \tag{7.4}$$

The traditional concept of density (number of individuals by square metre, i.e., $N/2$) for the measurement region poses a problem due to the small number of individuals involved when conducting an analysis in frames. Following Chattaraj Chattaraj et al. (2009), we determine a way to render this continuous by computing the information in seconds and to measure the number of individuals found per frame according to the video frame rate (in this case 30 FPS), and we compute the mean number of individuals presented in the ROI per second . In turn, densities are measured as values collected per second.

The following section discusses FD data obtained, i.e., speeds and densities, from the three countries. We also investigate the issue of personal space. We adopt the above described hypothesis (Jacques et al., 2007) to approximate personal space using a Voronoi Diagram (VD) (Aurenhammer, 1991). As our pedestrian tracking could not be applied to find out the order of pedestrians in the video (we do not know the order in which the pedestrians were tracked, e.g. $i$ and $i+1$), we use the output of the Voronoi Diagram to compute the neighbour of each individual (pedestrians in front and behind) to calculate distances between each pedestrian and his/her predecessor, which are defined as personal distance in this work (the magnitude of the vector from individual $i$ to his/her predecessor $i-1$).

We illustrate personal space (VD polygons) in Figure 7.5 for scenarios captured from Brazil ($N$ = 15 and $N$ = 20 (top) and $N$ = 25 and $N$ = 34 (bottom)). VD polygons shown in Figure 7.5 provide a visual account of pedestrian distributions observed in the corridor. Small areas (cool colours in the figure) represent the closest pedestrians while larger areas (warm colours in the figure) denote pedestrians that are more distant from one another.

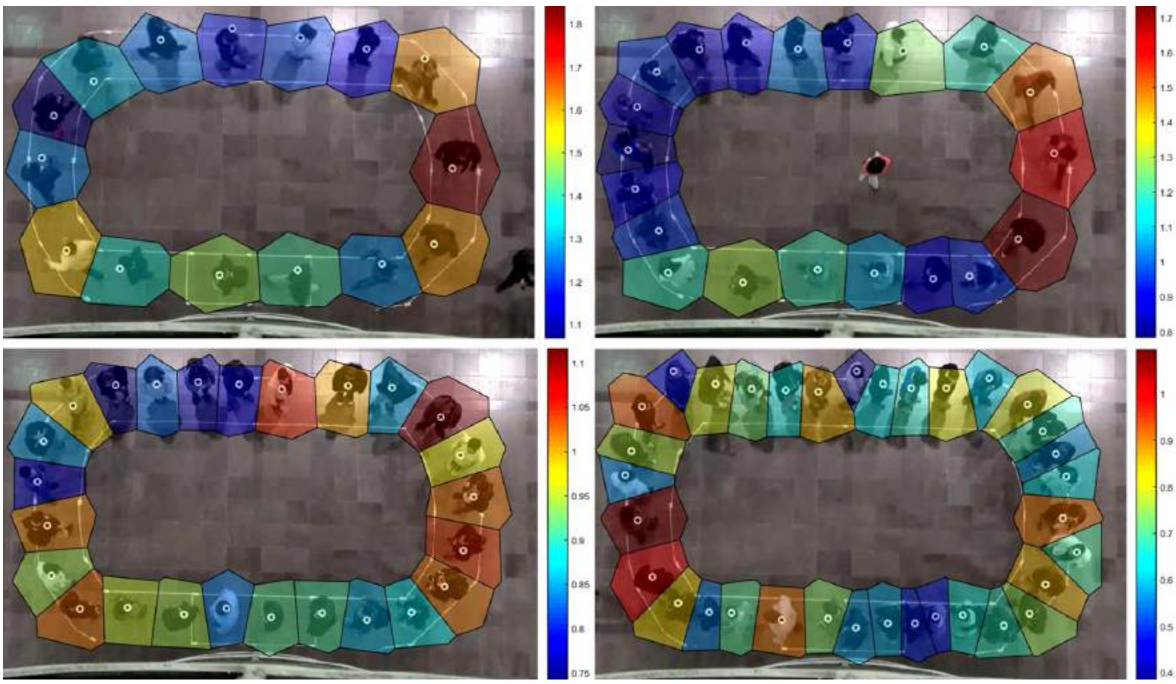


Figure 7.5: **Voronoi Diagram computed from a video sequence**: Voronoi Diagram showing, respectively, $N$ = 15 and $N$ = 20 (top) and $N$ = 25 and $N$ = 34 (bottom) individuals in Brazil. Values shown on the right denote the personal space (VD polygon area) in metres.

## 7.2 Fundamental Diagram Analysis

In this section, we first compare our results with those obtained by Chattaraj et al. (2009). The authors found that at low densities pedestrians from India and Germany behave in a similar way in terms of speed while at high densities certain differences are found between the countries, i.e., speeds vary more significantly. Figure 7.6 shows a box plot of speed distributions for the three countries and for all values of $N$ measured through the FD experiment. It is possible to observe that at low densities pedestrians from different cultures behave in a similar way with regard to speed variations (values vary less in relation to the mean). As densities increase, variation in speed also increases. As is shown in Chattaraj et al. (2009), Indians move at slower speeds than Germans at high densities. Brazilians travel at the highest speeds for all population sizes.

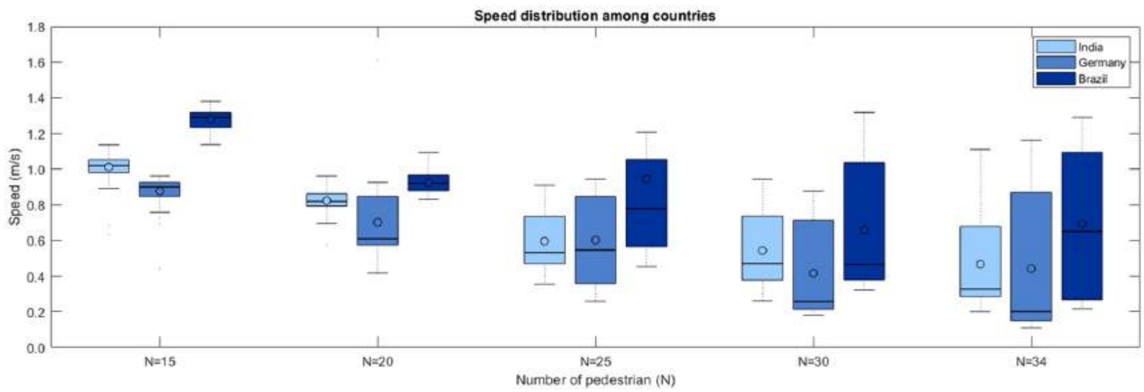



Figure 7.6: **Speed distributions**: speeds observed from the experiment.

Figure 7.7 shows an analysis of the Probability Distribution Function (PDF) applied for the observed personal speeds measured for the three countries. PDF function is a statistical expression that defines a probability distribution for a continuous random variable. When the PDF is graphically portrayed, the area under the curve will indicate the interval in which the variable will fall. Figure 7.7 shows the probability of distributions measured for each observed speed at an interval of $[0; 1.6]$ $m/s$. Individuals from Brazil present a higher probability of travelling at higher speeds than Indians and Germans. It is also true that as the $N$ increases, the means of PDF curves for the countries tend to converge to a similar value as stated in Table 7.1 with regard to the standard deviation. Thus, although speeds vary more when $N = 34$, mean values of the different countries are more similar.

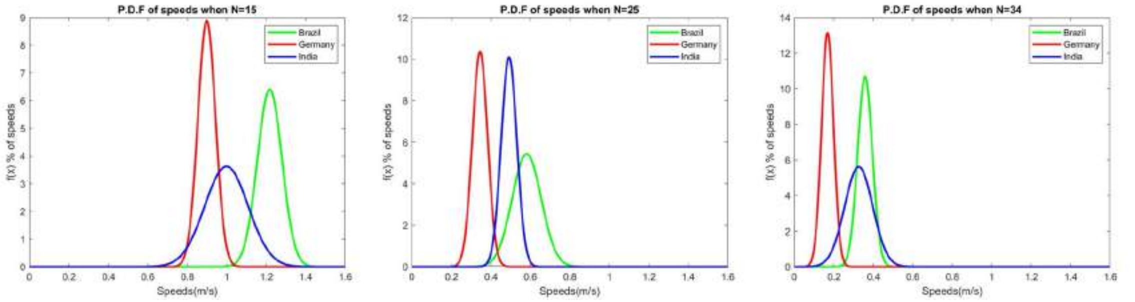



Figure 7.7: **Probability distribution function (P.D.F.)**: P.D.F. function determined from speeds measured during the experiment: (**left**) $N$ = 15, (**centre**) $N$ = 25 and (**right**) $N$ = 34.

Table 7.1: **Mean and standard deviation of P.D.F.**: mean and standard deviation values of the P.D.F. function observed for the countries with 3 $N$ values (15, 25 and 34).

| N | India | Germany | Brazil | Mean value | Std Deviation |
|---|---|---|---|---|---|
| $N$ = 15 | 0.9985 | 0.8985 | 1.2177 | 1.0382 | 0.1633 |
| $N$ = 25 | 0.4927 | 0.3453 | 0.5819 | 0.4733 | 0.1195 |
| $N$ = 34 | 0.3260 | 0.1684 | 0.3581 | 0.2842 | 0.1015 |

Similarly, Figure 7.8 shows an analysis of the PDF applied to observed personal distances of the three countries for an interval of [0; 2.5] metres. Distances from individuals measured in all countries become more similar as densities increase.

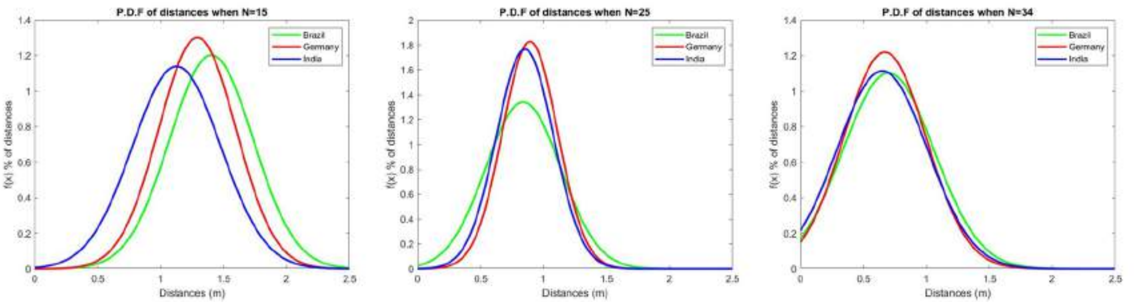



Figure 7.8: **Probability distribution function (PDF) from the distances**: P.D.F. function from distances in the experiment with, respectively: (**left**) $N$ = 15, (**middle**) $N$ = 25 and (**right**) $N$ = 34.

We sought to compare the three countries to find differences/similarities between them. We hypothesize that these differences are extractable through video analysis. Figures 7.9 and 7.10 show average speeds and densities as well as corresponding standard deviations for each cluster of the observed population (15, 20, 25, 30 and 34 individuals).

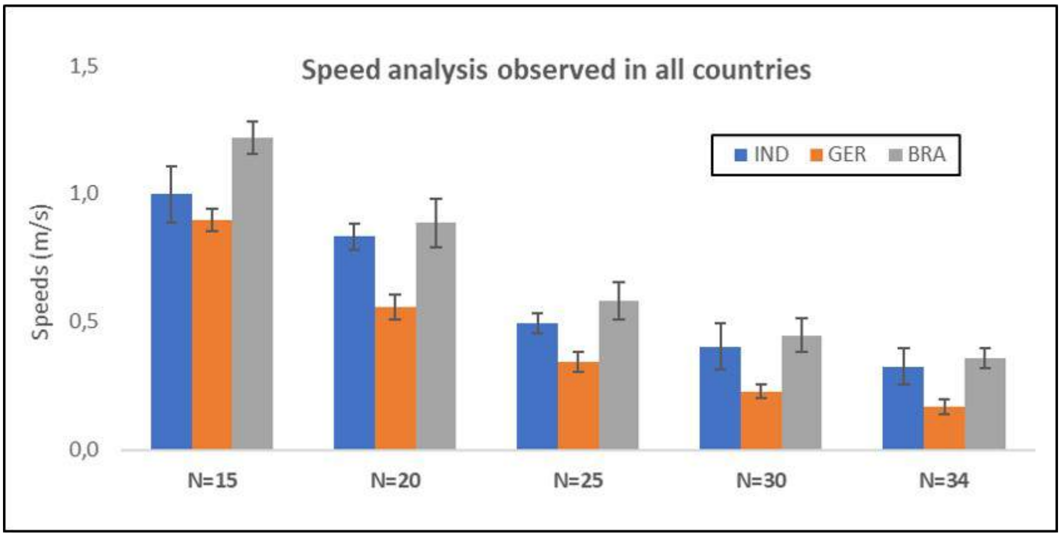



Figure 7.9: **Walking speed analysis**: walking speeds of the three countries and the observed standard deviation.

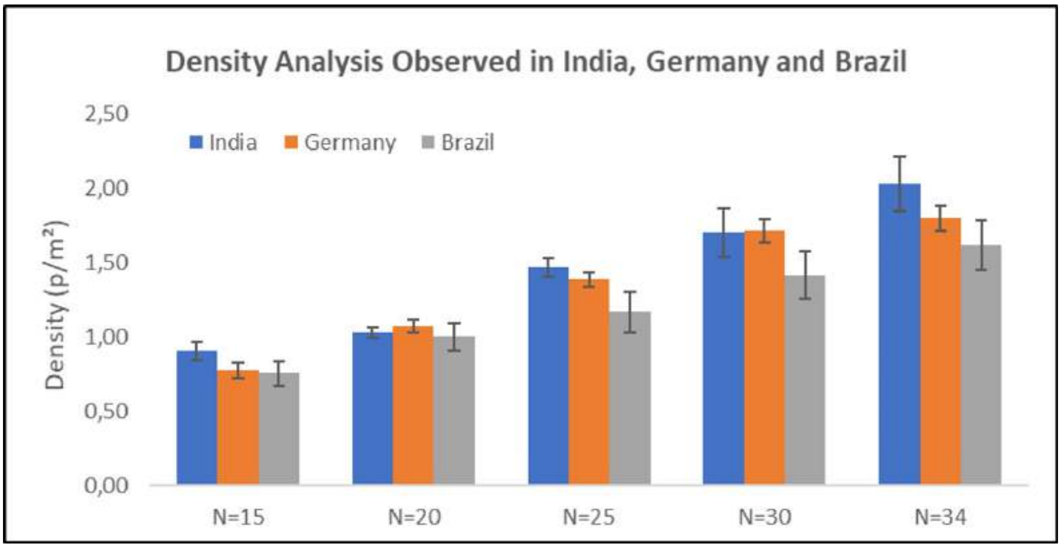



Figure 7.10: **Density analysis**: density of the three countries and the observed standard deviation.

As can be observed in Figure 7.9, Brazil presents higher speeds and even when speeds are more heavily impacted, i.e., at higher densities. On the other hand, Germany presents the lowest speeds of the three analysed countries. It is important to emphasize that the observed standard deviation in speed and density is higher for Brazil than for the other two countries. When visually observing videos collected from Brazil (videos collected from other countries are not available), one can see that some individuals adopt a "spring" behaviour, i.e., sometimes approaching from the person in front and sometimes maintaining distance. Concerning densities achieved by the populations (Figure 7.10), we note that India presents higher densities than Germany and Brazil. In addition, we find that standard deviations for the three countries increase with density as observed for speed.

In addition, we compute Pearson's correlation to find similarities and patterns of functions from the observed FD for the three countries. We measure the correlation between two sets of data with regard to the average speed, the speed standard deviation, the average density and the density standard deviation for each cluster of the observed population (15, 20, 25, 30 and 34 individuals). The results show that FDs for the three countries are most similar in average density (average Pearson's $r = 0.99$) followed by the density standard deviation (average Pearson's $r = 0.90$), the speed standard deviation (average Pearson's $r = 0.86$) and finally the average speed (average Pearson's $r = 0.79$), which still presents a high correlation. We thus show that FDs for the three countries are coherent and present some similarities.

This is coherent when we wish to find a pattern of FD for the different countries, and the results show high correlations and mainly in terms of density. Thus, some assumptions can be made based on our analyses:

1. There is a pattern of FD;

2. This observed pattern can be used to validate and/or calibrate simulations of virtual humans from games and movies (see *Fundamental Diagram Simulation* Section);

3. As FDs for India, Brazil and Germany are highly correlated, we assume that the tasks that individuals executed in our experiment were the same, and thus we attempt to measure the culturallity of this context as discussed in the next section; finally

4. We wish to investigate when the identified differences can be explained by cultural aspects and if they are observable when we investigate personal distance (see the Personal Space Analysis Section below).

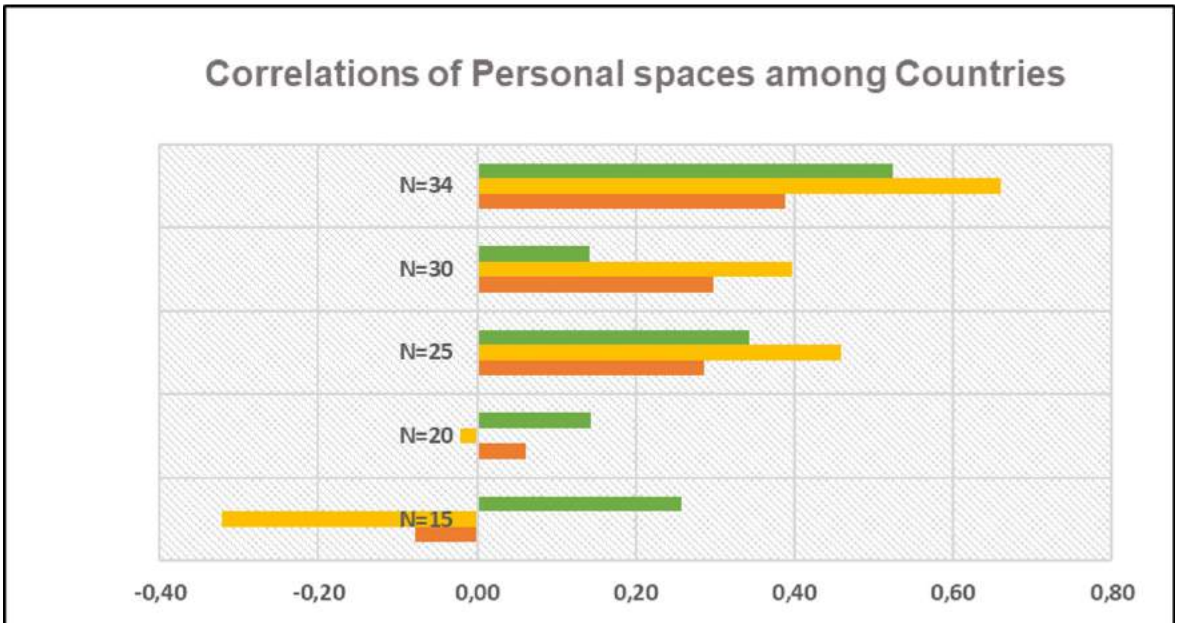
## 7.3 Personal Space Analysis

In this section, we present our personal space analysis of the FD experiment. We must first emphasize that as noted in the *Background* Section many studies have been conducted in this area in recent years. In most related work (Ballendat et al., 2010; Sorokowska et al., 2017), the distance between two individuals is defined as personal space. We also consider personal space, as we provide an approximation from the Voronoi Diagram (VD). However, for our analysis, we use the VD to determine the neighbourhood of individuals and to then extract those from which to compute distances. In this work we are interested in investigating cultural aspects of different populations regarding personal distance due to evidence showing that this can differ across countries (Sorokowska et al., 2017). As noted in the final section, assuming that the three analysed populations executed the same task, we extracted personal distances from FDs for our data analysis.

As for Brazilian population we had access to all of the experimental video coverage, we were able to track and determine the participants' positions in each frame. However, this was not true for the populations in India and Germany, for which we have information on the time at which each person entered and exited the analysed rectangle as sent by the authors (Chattaraj et al., 2009). We thus used this information to compute the speed of each person in the rectangle and the distance between two consecutive individuals in the rectangle (neighbours in the VD). We defined the distance from individual $i$ to the individual in front of him/her (individual $i+1$) as the personal space of $i$.

Correlations of distances between the three populations are shown in Figure 7.11. As can be easily observed, Pearson's correlations between the populations increase with density. Based on this, our hypothesis that high densities greatly impact personal behavioural expression makes sense as at high densities individuals act more as a mass and less as individuals (Vilanova et al., 2017), which ultimately affects their behaviours according to their own cultural backgrounds. This assumption is also coherent with one of the main works published on mass behaviour (Bon, 1986).

We also performed an evaluation of the walking speeds of individuals measured during the FD experiment. Therefore, the corresponding results are very similar to those of the distance analysis as shown in Figure 7.12. At lower densities India and Germany present a higher correlation ($r = 0.42$) while at higher densities



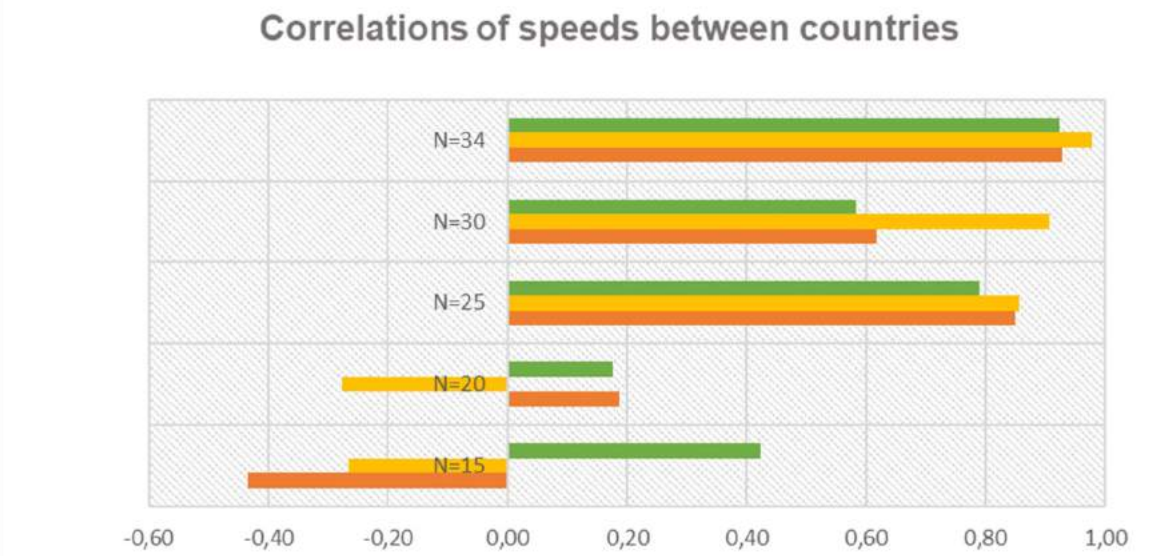

| | N=15 | N=20 | N=25 | N=30 | N=34 |
|---|---|---|---|---|---|
| IND-GER | 0,2579 | 0,1440 | 0,3428 | 0,1418 | 0,5243 |
| BRA-GER | -0,3216 | -0,0217 | 0,4587 | 0,3970 | 0,6596 |
| BRA-IND | -0,0788 | 0,0607 | 0,2856 | 0,2971 | 0,3888 |

Figure 7.11: **Correlations of personal space**: correlations measured for the examined countries.

all countries present very similar walking speeds. Again, one can speculate that cultural behaviours may be visible at lower and moderate densities, as at higher densities individuals behave as a mass restricted by reduced levels of free space. We must note that this correlation was computed based on all data observed each second and not based on means and standard deviations.

Thus, for this study, we assume that the difference in individual behaviour observed at the same densities and during the execution of the same task can be considered an expression of individuals' cultural backgrounds. For low to high densities, we observe similarities and differences in terms of personal distances and speeds (Figures 7.11 and 7.12). For 15 and 20 individuals measured in the experiments India and Germany are more similar with regard to personal distance and speed than those for Brazil. Indeed, in terms of personal distance, Brazil and Germany present more differentiated behaviours (even inversely proportional as shown by the negative correlation illustrated in Figure 7.11) when tested with 15 individuals.

Regarding individuals speeds, the Brazil vs. India and Brazil vs. Germany pairings present lower correlations for 15 individuals. We hypothesize that this varied behaviour observed at low densities can be explained by cultural backgrounds. As




| | N=15 | N=20 | N=25 | N=30 | N=34 |
|---|---|---|---|---|---|
| IND-GER | 0,4228 | 0,1757 | 0,7881 | 0,5843 | 0,9240 |
| BRA-GER | -0,2666 | -0,2763 | 0,8569 | 0,9061 | 0,9766 |
| BRA-IND | -0,4358 | 0,1863 | 0,8491 | 0,6183 | 0,9279 |

Figure 7.12: **Correlations of speed**: correlations measured for the examined populations.

stated above we visually found that Brazilian individuals adopted what we call '*spring behaviour*' in the FD experiment, which is an expression of their social behaviour.

Conversely, with increasing density, for experiments involving 25, 30 and 34 individuals for personal distances but mainly regarding speeds higher correlations are found between Brazil and Germany and an even higher correlation is found for 34 individuals. In any case, with higher densities all correlations are higher as well. This may indicate that when space is limited there is less room for individual behaviours to occur. According to our hypothesis, at low densities, India and Germany are more culturally similar. When free space is more limited, it seems that Brazil and Germany were more similar. Finally, at high densities individuals behave in a similar way, discarding their cultural behaviours and acting as a mass.

As we show in Figure 7.13, regarding the distance vs. speed (on the left) and distance vs. speed exponential relation (on the right) analyses, pedestrians from the different cultures seem to behave in a similar way at high densities, i.e., over shorter distances (please refer to distances around 0.5, for all countries, the speeds keep around $0.2 m/s$ ). Indian pedestrians, like Brazilians, tend to maintain shorter distances than German pedestrians, and they react less quickly to a change in predecessor speed compared to Germans, as shown in Chattaraj's work Chattaraj et al. (2009). Brazilian pedestrians seem to act in a similar way as Indians, but they maintain higher speeds over shorter distances (at higher densities). When we look at low densities, i.e., over higher distances (please refer to the distances around $1.4m$) the speeds of all countries are more different.

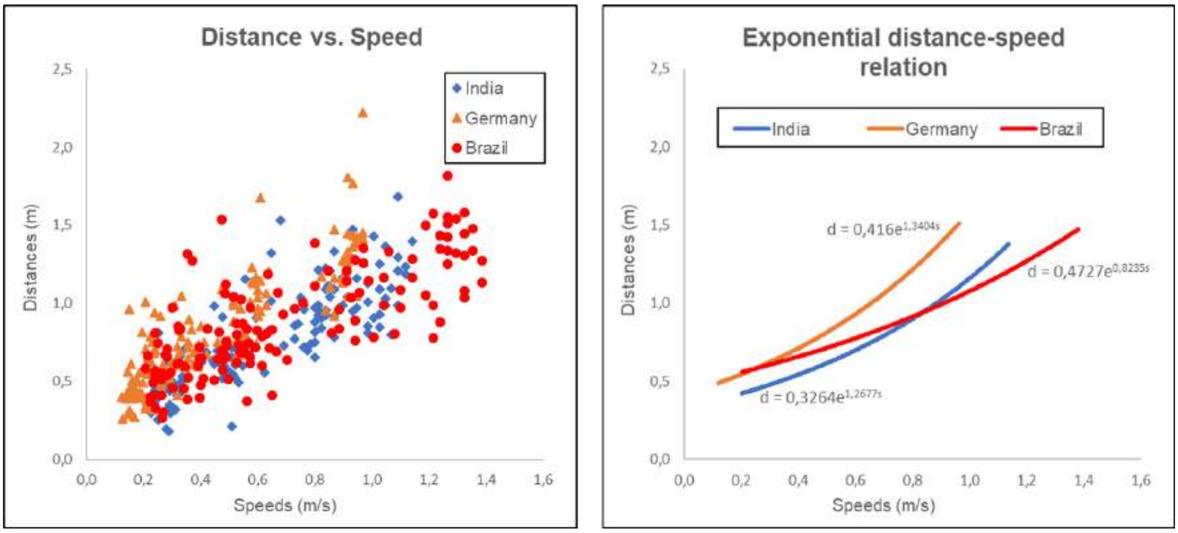



Figure 7.13: **Distance vs. speed and exponential distance-speed relation**: distance vs. speed (**left**) and fitted exponential distance-speed relations (**right**) analysis with the data from the three countries.

We also compared the preferred distances between individuals identified by Sorokowska et al. (2017) to results obtained from the experiment performed here. Next, we compared data for Germany[1] and Brazil. We must note that the videos from Germany are not exactly the same as those evaluated by Chattaraj et al. (2009) and do not involve all $N$ values, e.g., $N = 34$.

In Sorokowska's work, participant answers were given on a distance ($0-220cm$) scale anchored by two human-like figures labelled A and B. Participants were asked to imagine that they were Person A. Participants were then asked to rate how close Person B could approach for them to remain comfortable while having a conversation with Person B. Figure 7.14 shows a comparison of the four different FD scenarios involving 15, 20, 25 and 30 individuals (as illustrated in the German videos). In our study, we measured the distance person A maintained from person B positioned in front of him or her. For the comparison, from Sorokowska's study we

[1]We were granted access to videos for Germany with populations of 15, 20, 25 and 30 from authors of the PED experimental database available at http://ped.fz-juelich.de/db/

use an evaluation of acquaintances (individuals who are not close but not strangers) who are comparable to individuals examined in our experiment.

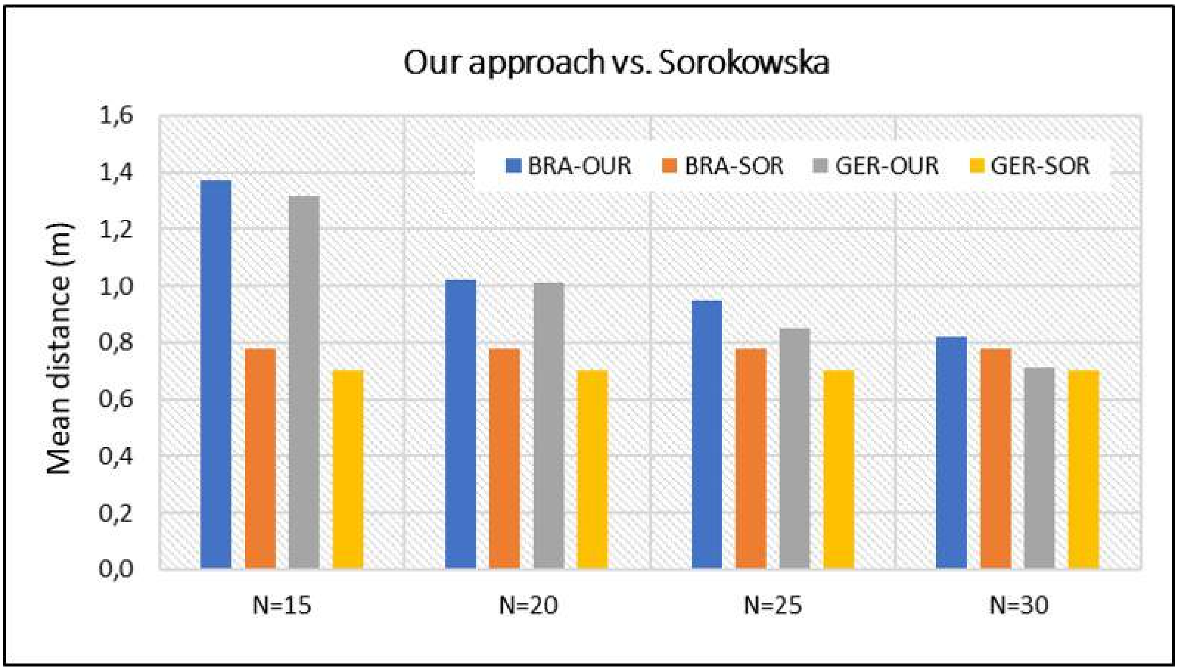



Figure 7.14: **Our approach vs. Sorokowska**: our results vs. Sorokowska et al. (2017) for different numbers of individuals examined.

As we show in Figure 7.14, while distances generated from our approach are higher than those identified by Sorokowska, proportions identified are similar across all scenarios. People from Brazil maintain more distance from others than individuals from Germany (according to our approach, for the $N$ = 15 configuration, individuals from Brazil position themselves 0.5 m further away from one another than they do in Germany while according to Sorokowska individuals from Brazil position themselves 0.8 m further away from one another). It is interesting that as the number of individuals increases, our values become more similar to Sorokowska's (when $N$ = 30 the values are quite similar). While the two experiments differ, we show in real conditions individuals actually behave according to the preferences described in Sorokowska's work.

In the next section we present an analysis regarding OCEAN and emotion traits in the fundamental diagram.

## 7.4 OCEAN and Emotion Analysis

In Chapter 5 we were interested about detecting personality and emotion traits in videos from different Countries, in public spaces. However, since the contexts are not exactly the same, even if the films are similar (see Section 4.3) e.g. people can be close or faraway from others not just because their personality but because the contexts they are evolving on or the relationship with spaces are not the same. Consequently, this research presents such difficult challenge to solve in order to evaluate/validate results. The ideal is to discard the spacial context in order to analyze only the people behavior, from different Countries, while executing the same task.

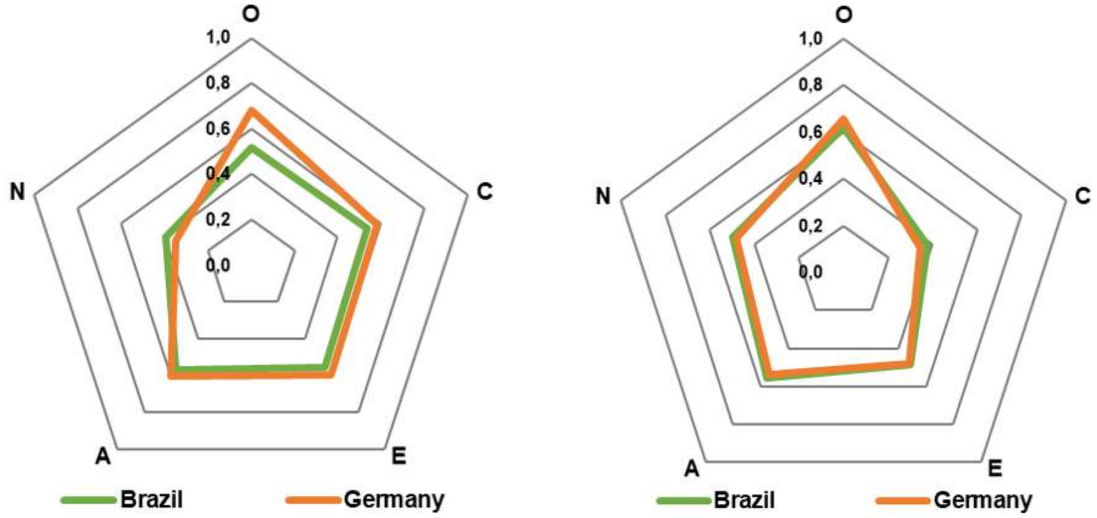
After tracked the individuals, we executed our method to extracted OCEAN (Section 5.2.1), Figure 7.15 shows the obtained OCEAN values from Brazil and Germany[2] in the experiments with extreme sizes of populations, i.e. $N = 15$ and $N = 34$ pedestrians. Based only on a visual inspection, we can easily perceive that OCEAN values from the both Countries are more similar when the density of people is higher in the experiment, as illustrated in Figure 7.15(b). It can indicate that people assumes group-level behavior instead of individual-level behavior caused by the higher density and the lack of free space. It agrees with several theories about mass behavior as discussed by Vilanova et al. (2017) and Bon (2012).

In addition, we mapped from OCEAN to emotion traits. In this analysis, we compute Pearson's correlation to find out the similarity between the two countries regarding the emotion aspects. Figure 7.16 shows the Pearson's correlation of the emotion values among the countries.

Considering the Figure 7.16, it is possible to see that as the density of pedestrians in the experiment increases, the correlation among the countries also increases.

We also included an investigation regarding the emotion and OCEAN impacted by the density of people. This was possible because the performed task is the same (i.e. individuals are walking in the same predefined environment) while only the number of people increases. Figure 7.17 shows how the emotion values obtained as the average of each video for all individuals, vary according to the den-

[2]We do not considered India in this analysis because we do not had access to the videos from the experiment performed in that country.




(a) $N$ = 15 pedestrians (b) $N$ = 34 pedestrians

Figure 7.15: **OCEAN analysis in a controlled experiment**: averaged OCEAN values observed in each country: (a) when $N$ = 15 and (b) when $N$ = 34 pedestrians.

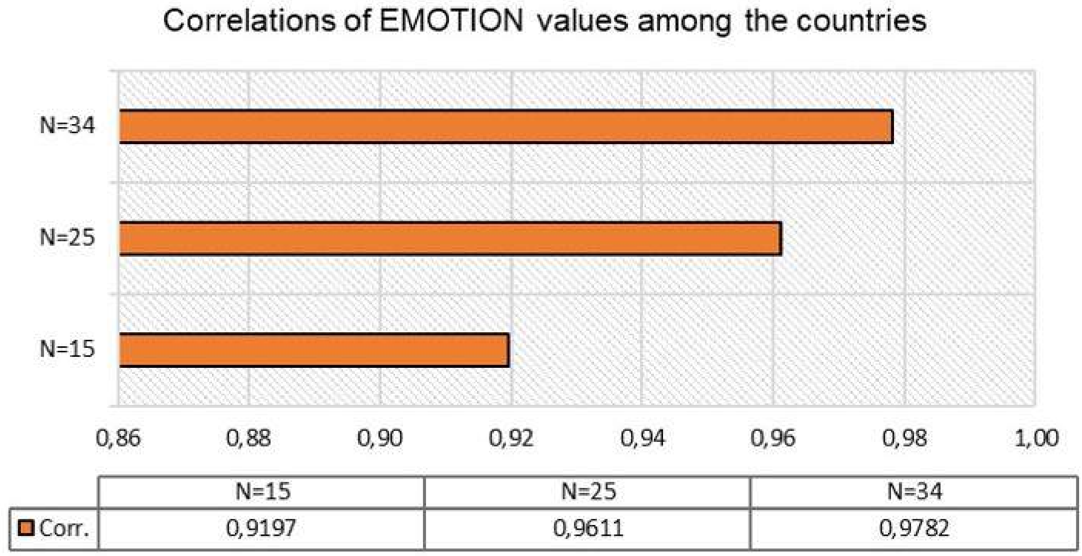



Figure 7.16: **Pearson's correlation of emotions**: values obtained in videos from Germany and Brazil, with different populations ($N$ = 15, $N$ = 25 and $N$ = 34).

sity of people in videos from Germany. It is interesting to see how the emotions Anger, Fear and Sadness increases proportionally as the density increases too.

On the other hand, Happiness emotion decreases proportionally as a function of observed density. Indeed, the data observed in Germany was in accordance to what was empirically expected in our hypothesis, i.e. the only positive emotion (H) decreases as the density increases. However, the computed emotion for Brazil

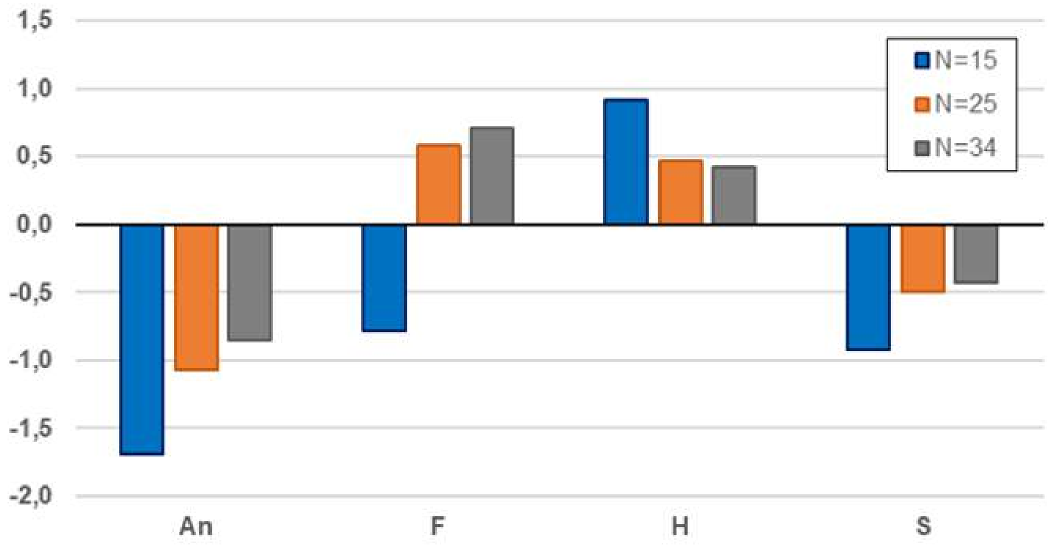



Figure 7.17: **Impact of emotion on density from Germany**: emotion of each density ($N$ = 15, $N$ = 25 and $N$ = 34).

(illustrated in Figure 7.18) was not in accordance to what we expected as in Germany. One possible explanation is that Brazilian people were colleagues/friends while in Germany, they were related by Chattaraj et al. (2009) as volunteers to the experiment.

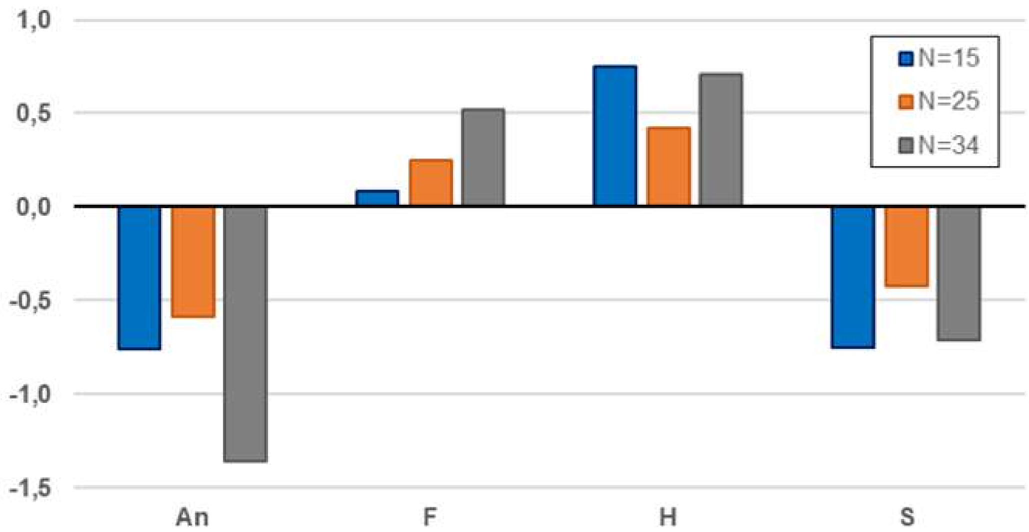



Figure 7.18: **Impact of emotion on density from Brazil**: emotion of each density ($N$ = 15, $N$ = 25 and $N$ = 34) from Brazil.

Even if the emotions are not impacted as a function of the density as presented in Germany graphic (Figure 7.17), it is interesting to see that for emotions like Anger and Sadness, Brazil and Germany presented negative average values, while Fear and Happiness presented positive average values.

Although we did not find any information about emotion and personality detection in videos from different Countries related in literature, these results seem promising in order to understand people personality, emotion and cultural differences in video sequences.

In the next section we use different simulation environments to reproduce patterns of FD in crowds.

## 7.5 Fundamental Diagram Simulation

Finally, in this section we investigate the behaviour of virtual individuals observed from two known crowd simulators, which are briefly presented in the following sections[3]. Our goal is to compare the results of simulations with patterns observed from FDs for different cultures. We in turn show how virtual crowds can be applied to games and movies.

### 7.5.1 Simulation Environments

We use two different simulations environments well known in crowd simulation area: BioCrowds (Bicho et al., 2012) and ORCA (Berg et al., 2009), both described in next sections.

#### BioCrowds

One agent in the environment perceives a set of markers (dots) on the ground (described through space subdivision) within its observational radius and moves forward to its goal based on such markers (unoccupied and closest to the agent than any other one). This is the main feature of the BioCrowds simulator (Bicho et al., 2012), which supports main behaviours observed from crowd simulations

[3]This research was supported by colleagues Estêvão Testa (who performed the experiments in *ORCA*) and Paulo Knob (who performed the experiments in BioCrowds).

(e.g., lanes and arcs formation) as also emergent from other crowd simulators (Helbing and Johansson, 2011; Van den Berg et al., 2008). Due to the main functions of BioCrowds method, obstacles are easy to represent as zones without any markers via space discretization.

As output, BioCrowds measures the position of each agent at each frame $X_j^f$. For more information on this method, please refer to Bicho et al. (2012). In this work we simulated an FD using BioCrowds (see Figure 7.19) for the same populations tested using a similar environmental setup, as described in Chattaraj et al. (2009).

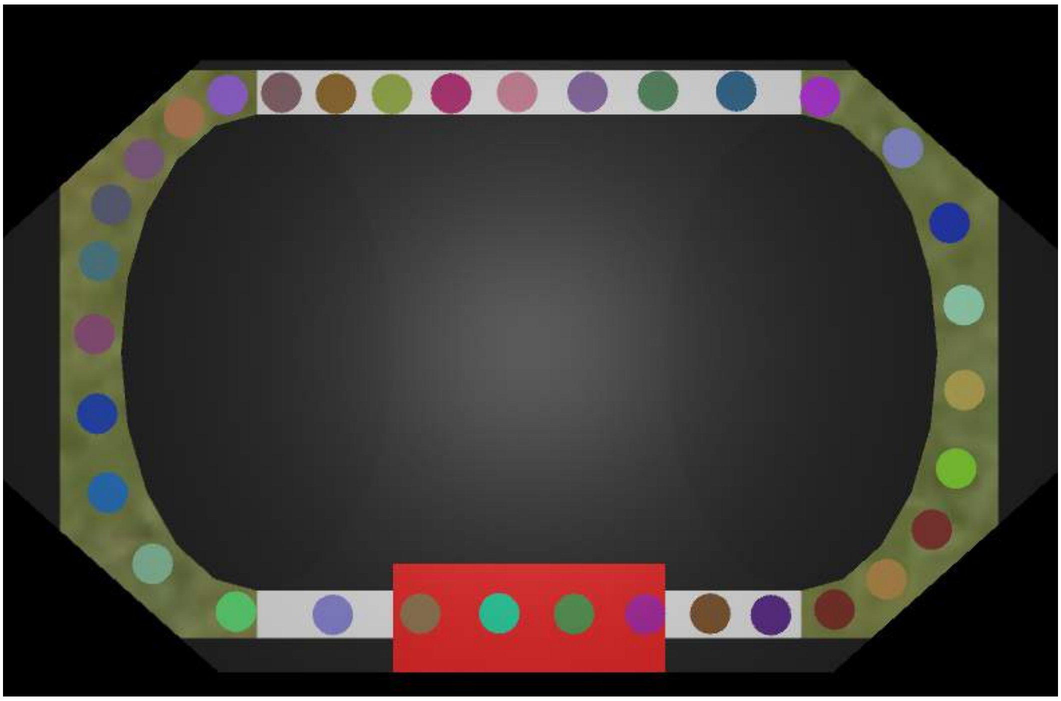


Figure 7.19: **Virtual agents simulated by BioCrowds**: we illustrate 34 virtual agents evolving during the FD experiment simulated by BioCrowds.

Optimal Reciprocal Collision Avoidance (*ORCA*)

*ORCA* prevent collisions from occurring between multiple agents as proposed by Berg et al. (2009). The objective of the method is to find the optimal velocity for each agent such that each agent manages to move through the environment without colliding and where each agent considers one another's velocity. The method was developed for use in robotics to prevent robots from colliding and as one of its greatest features it runs at $\mathcal{O}(n)$ without needing to maintain communication between robots, as they only need to perceive one another's positions and velocities. *ORCA* allows each robot to be free of collisions for at least a fixed amount of time assuming that each robot uses it as a collision avoidance protocol.

Agents are represented with a simple circle shape and are holonomic, i.e., they can move in any direction without needing to turn around. A possible velocity for agent $A$ is computed by finding permitted velocities $V_A$ between agent $A$ and each other agent. To do this, the velocity obstacle($VO$) measured between $A$ and each other agent $B$ for time $T$ is determined. In other words, $VO$ measures all relative velocities of $A$ with respect to $B$ that will result in a collision between them during time $T$. As performed in BioCrowds, we simulated an FD using *ORCA* (see Figure 7.20). Corresponding results are illustrated in the right-hand image of Figure 7.21. The results obtained by BioCrowds are illustrated in the left image of Figure 7.21.

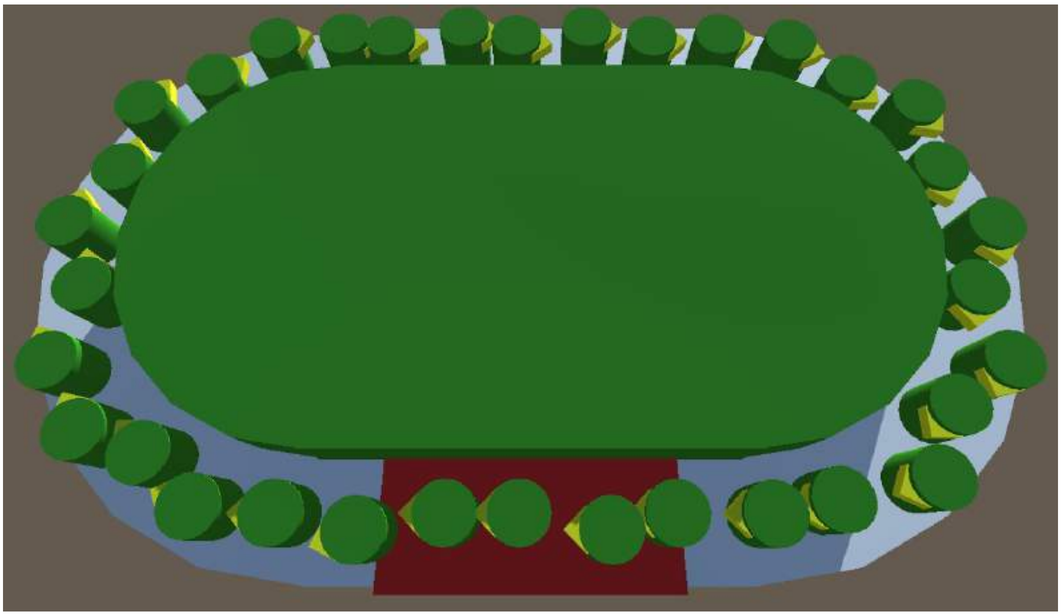


Figure 7.20: **Virtual agents simulated by** *ORCA*: we illustrate 34 virtual agents evolving during the FD experiment simulated by *ORCA*.

Simulation Setup

In both simulators we first modelled the $3D$ environment to have exactly the same dimensions as those of the experiment performed in real life. For BioCrowds and ORCA we manually instantiated the agents using Unity and we defined agent goals. Goals denote the four corners of the elliptic space and change (as shown in the original models) for the following goal once each one is achieved. For both implementations we change the number of agents based on scenarios of the real FD.

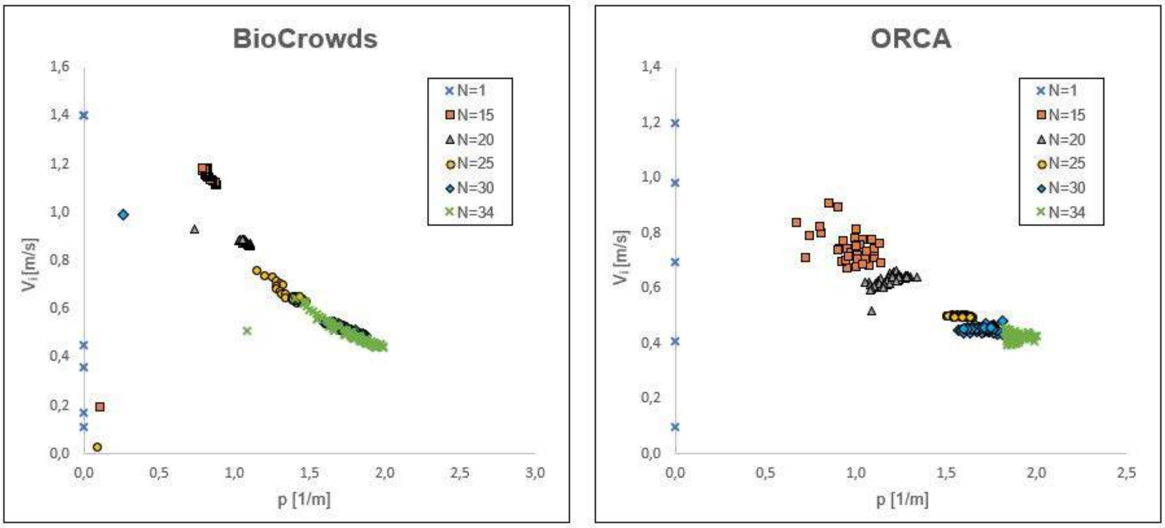
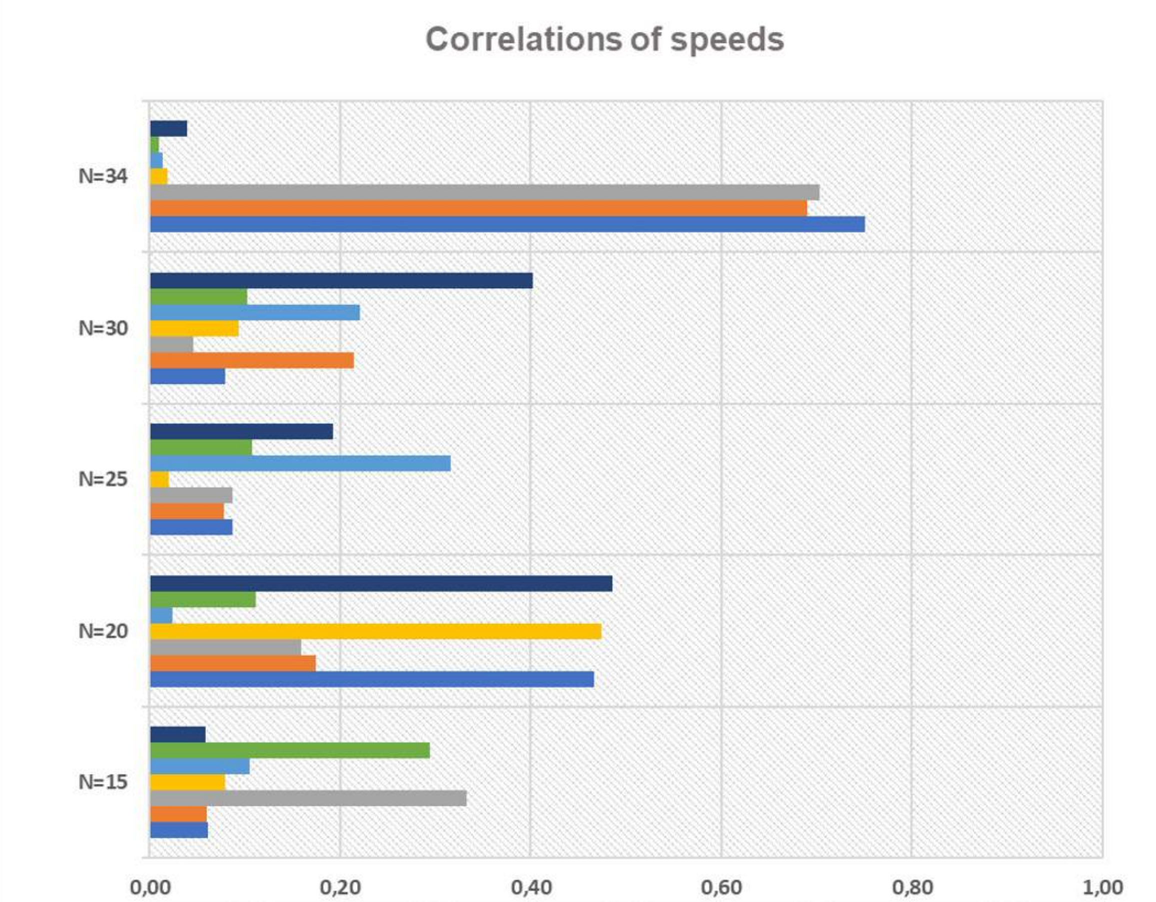



Figure 7.21: **Fundamental Diagram from simulation environments**: FDs obtained from BioCrowds (left plot) and *ORCA* (right plot) simulation environments. The vertical axis denotes variations in speed and the horizontal axis denotes density values.

In both simulators, agents could overcome the one in front of it when there was enough space to achieve their goals. Thus, to guarantee the order of agents, we applied an additional procedure to prevent this. The speed was defined in BioCrowds as the maximum speed, i.e., $1.2m/s$. However, in *ORCA* we changed the desired speed as a function of density (from $1.2m/s$ to $0.5m/s$ depending on $N$ for 1 to 34 agents). This is the case because the method predicts the future movements of other agents to find an optimal velocity that can generate unrealistic movements for denser crowds. We are aware that improvements have been made to this method regarding such characteristics as proposed by Berg et al. (2009), but these improvements were not tested in this work.

### 7.5.2 Simulation Results

We also computed correlations between FDs for real individuals in India, Germany and Brazil and compared them to *ORCA* and BioCrowds simulation results. Figure 7.22, which shows the obtained results, illustrates that the correlations are no greater than $r = 0.50$ for all tested populations (15, 20, 25 and 30 individuals). Only at higher densities (34 individuals) we can observe the higher correlation found from BioCrowds and for the analysed countries. We hypothesize that even in the BioCrowds simulator "cultural aspects” (in this case denoting the “simulation of virtual humans in BioCrowds”) are lost when free space is limited. This is an interesting result for BioCrowds researchers, as our study of cultural aspects of FDs can be applied to evaluate and improve crowd simulators.




| | N=15 | N=20 | N=25 | N=30 | N=34 |
|---|---|---|---|---|---|
| ORCA-BIOC | 0,06 | 0,49 | 0,19 | 0,40 | 0,04 |
| ORCA-BRA | 0,29 | 0,11 | 0,11 | 0,10 | 0,01 |
| ORCA-GER | 0,11 | 0,02 | 0,32 | 0,22 | 0,01 |
| ORCA-IND | 0,08 | 0,47 | 0,02 | 0,09 | 0,02 |
| BIOC-BRA | 0,33 | 0,16 | 0,09 | 0,05 | 0,70 |
| BIOC-GER | 0,06 | 0,17 | 0,08 | 0,21 | 0,69 |
| BIOC-IND | 0,06 | 0,47 | 0,09 | 0,08 | 0,75 |

Figure 7.22: **Correlations between FDs in real and simulated environments**: correlations between FDs for real individuals in India, Germany and Brazil in comparison to *ORCA* and BioCrowds simulation tools.

## 7.6 Chapter Remarks

In this chapter, we discuss differences in cultural features of groups of individuals observed in video sequences. We initially expected to find cross-cultural manifestations to vary from video sequences. However, as one important aspect to be considered in behaviour analysis is the context and environment in which individuals behave, we decided to exclude such variations by fixing tasks that the tested populations were required to execute. This is why we used Fundamental Diagrams proposed by Chattaraj et al. (2009).

We obtained information for India and videos from Germany and conducted our experiment in Brazil using the same environmental setup. All analysis results presented in this paper are related to these experiments, for which we hypothesized that by fixing the environmental setup and individual tasks, we could evaluate cultural variations in individual behaviour.

From this analysis we make contributions to pattern recognition focused on video understanding and specifically in regards to cultural features of groups of individuals as follows:

- From our FD analysis, we observe decreasing walking speeds with increasing density. This validates FD usage for the three populations. While India and Germany were already evaluated in Chattaraj et al. (2009), this paper contributes with data for Brazil. We must note that Pearson's correlations obtained for the countries are higher at $r$ = 0.79 (average speed) to $r$ = 0.99 (average density). In addition, we found coherent results with higher speeds always observed for Brazil and with higher densities observed for India all tested population sizes (see Figures 7.9 and 7.10).

- From our personal distance analysis, we found that as the density of individuals increases, individuals are more homogeneous, as shown by the computed Pearson's correlation in given Figures 7.11 and 7.12. This shows that individuals exhibit mass behaviours instead of behaving individually according to their cultural background or personalities. This serves as interesting and concrete proof of several theories on mass behaviour as discussed by Vilanova et al. (2017); Bon (2012).

- The obtained results are in accordance with work conducted in this area Chattaraj et al. (2009) in that at low densities individuals vary less in speed than at higher densities and in other ways as described in the text.
- We think that we have found strong evidence of the fact that the cultural behaviours of individuals can be observed at low densities, and from this, we can define better means to observed events in video sequences. Such cultural behaviour is heavily impacted by the speeds at which individuals travel and by distances maintained from others in social spaces. Temporary changes in such behaviours are also important and will be investigated in future work.
- Finally, the crowd simulation results present "cultural findings", e.g., main characteristics of the proposed solution, that disappear at high densities. This is important to consider when using computational applications (e.g., hazardous simulations).

The results of this research are relevant to three practical areas related to surveillance and computer simulations:

- Our work shows that in dense crowds, cultural and individual differences are likely to disappear due to limited free space. Thus, the integration of cultural aspects of surveillance systems for measuring pedestrian actions and behaviours should be performed only at low and moderate densities. In such cases, speeds and distances are important to consider in evaluating cultural aspects.
- When simulating cultural crowds for games or movies as in Dihl et al. (2017), one should focus on parameter variations (speeds and distances) observed at low and moderate densities of individuals. It can be of benefit to know that at a certain density (high), cultural factors are not as relevant as they are at low and medium densities.
- We must note that when considering safety systems, for cases of crowd evacuation, our research can contribute in the sense that in dense crowds, individuals may respond in a homogeneous way, as cultural and individual factors do not seem to interfere in such situations. However, at low and moderate densities, individual differences can produce different results. Of course, in the specific case of crowd behaviours, other variables are also important (e.g., individual training, previous knowledge of the environment, group behaviour).

We find this to be an interesting topic of research that should be explored to consider cultural features of crowd evacuation.

Regarding personalities and emotions, an important challenge in this area is the comparison with real life data. In this work we successfully compared our results with OCEAN from two specific countries (Brazil and Germany) present in Psychological literature. As one particular aspect to be considered in behavior analysis is the context and environment in which individuals behave, we decided to exclude such variations by fixing tasks that the tested populations were required to execute. This is why we used Fundamental Diagram for pedestrians. To do that, we performed a full experiment in Brazil to serve as benchmark to our research.

The next chapter presents the final considerations of this thesis, listing the main contributions and the future directions of this thesis.

# 8. FINAL REMARKS

The study of groups, crowds and their dynamics is not only essential for understanding pedestrians behaviour, but also for understanding communities, cultures and society itself (Forsyth, 2018). In this sense, the present thesis investigate pedestrian and crowds characteristics, regarding personalities, emotions and cultural aspects. The main goal of this research was to propose a model that could detect pedestrian features in crowds from video sequences, allowing the comparison and analysis of pedestrian characteristics from different crowds or countries.

According to Forsyth (2018), because each society is unique in its traditions and culture, the groups within any given culture may display unique interpersonal behaviours and characteristics. The study of crowds in Computer Science is mainly focused on entertainment and security. However, there are not many research addressing the cultural aspects of these crowds. We are interested in finding characteristics of pedestrians in crowds that are able to describe them and, in a certain way, differentiate the population from a crowd to another.

Extracting characteristics from real pedestrians and crowds, benefiting several other areas of knowledge, such as: architecture and design (planning spaces to maximize pedestrian and group-environment fit and design of spaces for groups); security and surveillance (design of evacuation plans considering characteristics of the crowds and detection of abnormal events); entertainment (more realistic crowds in movies and games reproducing characteristics from real pedestrians and crowds); social sciences (understanding of human behavior), among others.

In this context, we proposed the **Big-Four Geometrical Dimensions** or just **Big4GD**, a model containing a set of pedestrian and crowd features grouped into four dimensions: *I - Physical*, *II - Social*, *III - Personal and Emotional* and *IV - Cultural*. This four dimensions enclose characteristics in different levels, as pedestrian, group and crowd. This model is an approach to detect these dimensions at a specific moment in a certain physical space, from the geometric point of view and not a scientific tool for assessing cultural, personality or emotional profiles of population.

Based on geometric characteristics derived from pedestrian trajectories, we propose a way of characterizing pedestrians and groups of individuals in crowds, allowing the comparison of each other to find differences between one crowd and another. Based on a series of experiments, we were able to validate our model and verify that our approach succeeds in extracting the information from the crowds and their individuals.

In chapter 2 (*Background*) we presented a series of concepts involving cultural, personality and emotion aspects, as well as their main classification models used in this work. In addition, several concepts involving crowds were presented, such as proxemics, which treats the interpersonal spaces around the individuals when they are interacting with others in a crowd and Fundamental Diagrams, used to analyze the relationship between density, velocity, and flow of individuals in a crowd, which can be influenced by cultural and personal aspects.

Chapter 3 (*A State-of-the-art Review*) presented several related work to what is being proposed in this thesis. As the main differences, we use in addition to controlled experiments, Computer Vision techniques in spontaneous videos of crowds and this thesis aims to detect cultural aspects and personality traits of pedestrians based solely on geometric characteristics instead of creating crowds and behaviors in simulated environments using cultural aspects and personalities as inputs.

In Chapter 4 (*First and Second Dimensions: Data Extraction, Crowd Types and Video Similarity*) we presented and discussed the results obtained in the initial steps of the geometrical dimensions model: *I - Physical* and *II - Social*. We also presented in this chapter a metric to find out similarity between crowds in videos. It is important because when comparing two crowds of different countries to find cultural differences, it is necessary that they have similar characteristics, otherwise there would probably be differences masked by the features of the places where the videos were recorded.

Chapter 5 (*Third Dimension: Detection of Personality and Emotion Traits*) was responsible for present the third dimension of Big4GD: *III - Personal and Emotional*. We propose to detect OCEAN personality traits and compared with data from different countries existent in the literature. In addition, based on OCEAN we compute four traits of emotion defined in OCC model. In Chapter 6 (*Fourth Dimension: Detection of cultural aspects in crowds*) we presented the fourth dimension of Big4GD: *IV - Cultural*. In this chapter we presented how we proceed to characterize groups of pedestrians in video sequences and detect some cultural aspects inspired in cultural dimensions proposed by Hofstede (2001).

Chapter 7 (*Fundamental Diagram Analysis*) described an experiment we performed involving Fundamental Diagrams. We analyzed cross-cultural manifestations in video sequences from three countries: Brazil, Germany and India. As one important aspect to be considered in behaviour analysis is the context and environment in which individuals behave, we decided to exclude such variations by fixing tasks that the tested populations were required to execute. This is why we used Fundamental Diagrams proposed by Chattaraj et al. (2009). Regarding speeds, we observed decreasing walking speeds with increasing density. From our personal distance analysis, we found that as the density of individuals increases, individuals are more homogeneous, in accordance with the theories on mass behaviour as discussed by Vilanova et al. (2017) and Bon (2012).

A big challenge in the subject of research of this thesis is the comparison with real life data. In this way, regarding personality and cultural aspects, we successfully compared our results with Psychology literature, where several studies aimed to analysis human behavior. It is important to notice that, even if the literature measured these dimensions by considering a different type of information (subjective responses of individuals collected through questionnaires), the results obtained from our approach using geometrical information indicate that our model generates coherent information when compared to data provided in available literature, as shown in various analysis in this manuscript.

Concerning perception of extracted data, we performed an experiment to evaluate if people can perceive different emotions and personalities performed by pedestrians in crowds. The results were satisfactory, even without explaining to the participants the concepts of each emotion or personality and how they were calculated in our approach (considering the geometric characteristics), in all the cases, most of the participants perceived the personality and emotion that the agent was expressing in the video, in accordance with our approach.

In addition, as one particular aspect to be considered in behavior analysis is the context and environment in which individuals behave, we performed an experiment (Fundamental Diagram) excluding such variations by fixing tasks that the tested populations were required to execute. In this controlled experiment, we have found strong evidence of the fact that the cultural behaviours of individuals can be observed at low and medium densities, i.e., in high densities individuals exhibit mass behaviours instead of behaving individually according to their cultural background or personalities. This serves as interesting and concrete proof of several theories on mass behaviour as discussed by Bon (2012) and Vilanova et al. (2017).

Besides to the geometric dimensions model proposed in this thesis, our research generated other important contributions to this area:

i) *Cultural Crowds* dataset: a public dataset of videos with crowds of different countries. Together these videos are files with tracking, pedestrian and crowd features and geometric dimensions information; and

ii) *GeoMind*: we developed a software for the detection and analysis of geometric dimensions **Big4GD** in video sequences.

Some future directions should be focused on investigating other models of pedestrians and crowd behaviours in the literature of psychology and behavioral sciences. In addition, we summarize a set of future work to recommend in order to improve the proposed approach and add new possibilities to our research:

i) *To evaluate the human perception about HCD*: in this thesis we conducted an experiment to evaluate the participants' perception about the personalities and emotions on virtual humans simulation. As a future work, we intend to carry out a similar experiment, evaluating the participants' perception about the Cultural Dimensions of Hofstede and other geometric characteristics not covered in the experiment already carried out;

ii) *To perform an experiment with pedestrians regarding Big-Five*: for validation purposes, we used the Big-Five dimension values present in the literature, at the country level (being an average of the personality dimensions of the members of that country) and a perception analysis experiment of some dimensions at the pedestrian level. We intend, in partnership with psychologists, to apply the NEO PI-R personality test to participants, then divide these participants into groups with high and low scores of OCEAN dimensions, and finally, ask these groups to perform some behavioral tasks in the same space, allowing us to evaluate the effectiveness of our model;

iii) *To use Machine Learning techniques to detect Big-Five dimensions*: Once performing more experiments involving the application of the personality test *NEO PI-R* in people and then filming them, we want to investigate the possibility of using machine learning techniques to perform the detection of each of Big-Five dimensions;

iv) *To increase the video dataset*: we built a video dataset names *Cultural Crowds*. This dataset contains videos from crowds with different characteristics. We intend to increase this dataset in both aspects, number of countries and the number of videos from each of them. One of the major difficulties of this work was to find a suitable set of videos to perform the experiments;

v) *To add new functionalities to GeoMind*: as one of the contributions of this thesis, we developed a software called *GeoMind*. This software detects the Big-Four Geometrical Dimensions from video sequences. We intend to add new functionalities, for example the ability to perform comparisons at different levels, such as comparing the geometrical dimensions of crowds from different countries or comparing all the features from two pedestrians.

In the Appendix A (*Video Analysis Dataset and Applications*) we present some important contributions developed in this thesis: i) *Cultural Crowds* dataset, a video dataset with crowds from various countries and Fundamental Diagram videos from Brazil; ii) *GeoMind*, a software to detect and analyze the Big-Four Geometrical Dimensions proposed in this thesis and iii) a viewer using a simulation environment to visualize the geometrical features of pedestrians.

As the text production associated to this thesis, we presented in Appendix B (*Publications*) the relation of publications obtained during the development of this research. In addition, some ongoing work are being developed or already submitted.

As a final word, we want to mention that after all the performed experiments and analysis, some interesting finds could be highlighted in this thesis:

i) Our main hypothesis that pedestrians behave according to their internal intrinsic characteristics such as personality, emotion and cultural aspects can be observed in terms of physical and geometrical behaviours; and

ii) In terms of perceptions, people can perceive this physical/geometrical manifestation of the pedestrians behavior.

# References

Abele, A. E. and Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of personality and social psychology*, 93:751–763.

Alameda-Pineda, X., Ricci, E., and Sebe, N. (2018). *Multimodal Behavior Analysis in the Wild: Advances and Challenges*. Elsevier Science, London, UK.

Albeverio, S., Andrey, D., Giordano, P., and Vancheri, A. (2007). *The Dynamics of Complex Urban Systems: An Interdisciplinary Approach*. Physica-Verlag Heidelberg, Mendrisio, SW.

Alcantara, M., Testa, E., da Silva, G. L., Favaretto, R., Dihl, L., and Musse, S. (2016). Generating background population for games based on real video sequences. In *Brazilian Symposium on Games and Digital Entertainment (SBGames)*, pages 66–72, São Paulo, SP. SBGames.

Allain, P., Courty, N., and Corpetti, T. (2014). Optimal crowd editing. *Graphical Models*, 76:1–16.

Andersen, M. and Taylor, H. (2007). *Sociology: understanding a diverse society, updated*. Cengage Learning, Belmont, CA.

Aurenhammer, F. (1991). Voronoi diagrams: a survey of a fundamental geometric data structure. *ACM Computing Surveys*, 23:345–405.

Ballendat, T., Marquardt, N., and Greenberg, S. (2010). Proxemic interaction: designing for a proximity and orientation-aware environment. In *International Conference on Interactive Tabletops and Surfaces*, pages 121–130, New York, NY. ACM.

Bassi, B. (2006). *Computer Simulation of Crowd Dynamics and Destructive Crowd Behavior*. (honors scholar thesis), University of Connecticut, USA, Storrs, CT.

Berg, J., Guy, S. J., Lin, M. C., and Manocha, D. (2009). Reciprocal n-body collision avoidance. In *Robotics Research*, pages 3–19, Berlin, GE. Springer Berlin Heidelberg.

Berlonghi, A. E. (1995). Understanding and planning for different spectator crowds. *Safety Science*, 18:239–247.

Berry, J. W., Poorting, Y. H., Breugelmas, S. M., Chasiotis, A., and Sam, D. L. (2011). *Cross-cultural psychology: research and applications*. Cambridge University Press, Cambridge, NY.

Best, A., Narang, S., Curtis, S., and Manocha, D. (2014). Densesense: interactive crowd simulation using density-dependent filters. In *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 97–102, Switzerland, SW. Eurographics Association.

Bicho, A. L., Rodrigues, R. A., Musse, S. R., Jung, C. R., P., M., and Magalhães, L. P. (2012). Simulating crowds based on a space colonization algorithm. *Computers & Graphics*, 36:70–79.

Bins, J., Dihl, L. L., and Jung, C. R. (2013). Target tracking using multiple patches and weighted vector median filters. *Journal of Mathematical Imaging and Vision*, 45:293–307.

Bon, G. L. (1986). *A Psicologia das Multidões*. Presses Universitaires de France, Paris, FR.

Bon, G. L. (2012). *The Crowd: A Study of the Popular Mind*. Courier Corporation, Mineola, NY.

Cai, Z., Yu, Z. L., Liu, H., and Zhang, K. (2014). Counting people in crowded scenes by video analyzing. In *9th Conference on Industrial Electronics and Applications*, pages 1841–1845, Hangzhou, CN. IEEE.

Cao, S., Lian, L., Chen, M., Yao, M., Song, W., and Fang, Z. (2018). Investigation of difference of fundamental diagrams in pedestrian flow. *Physica A: Statistical Mechanics and its Applications*, 506:661–670.

Cao, S., Seyfried, A., Zhang, J., Holl, S., and Song, W. (2017). Fundamental diagrams for multidirectional pedestrian flows. *Journal of Statistical Mechanics: Theory and Experiment*, 2017:33404–33412.

Cattell, R. B. and Krug, S. E. (1986). The number of factors in the 16pf: a review of the evidence with special emphasis on methodological problems. *Educational and Psychological Measurement*, 46:509–522.

Chan, A. B. and Vasconcelos, N. (2009). Bayesian poisson regression for crowd counting. In *12th International Conference on Computer Vision (ICCV)*, pages 545–551, Kyoto, JP. IEEE.

Chandran, A. K., Loh, A. P., and Vadakkepat, P. (2015). Identifying social groups in pedestrian crowd videos. In *Eighth International Conference on Advances in Pattern Recognition (ICAPR)*, pages 1–6, Kolkata, IN. IEEE.

Chattaraj, U., Seyfried, A., and Chakroborty, P. (2009). Comparison of pedestrian fundamental diagram across cultures. *Advances in Complex Systems*, 12:393–405.

Colque, R. V. H. M., Júnior, C. A. C., and Schwartz, W. R. (2015). Histograms of optical flow orientation and magnitude to detect anomalous events in videos. In *28th Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 126–133, Salvador, BR. IEEE.

Costa, P. and McCrae, R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI)*. Psychological Assessment Resources, New York, NY.

Costa, P. and McCrae, R. R. (2007). *NEO PI-R - Inventário de Personalidade NEO Revisado*. Vetor, São Paulo, SP.

Costa, P. T. and McCrae, R. R. (1989). *NEO PI/FFI manual supplement for use with the NEO personality inventory and the NEO Five-Factor Inventory*. Psychological Assessment Resources, New York, NY.

Davidich, M. and Koster, G. (2013). Predicting pedestrian flow: a methodology and a proof of concept based on real-life data. *Plos One*, 8:833–855.

Davies, A. C., Yin, J. H., and Velastin, S. A. (1995). Crowd monitoring using image processing. *Electronics & Communication Engineering Journal*, 7:37–47.

Davis, K. L. and Panksepp, J. (2011). The brain's emotional foundations of human personality and the affective neuroscience personality scales. *Neuroscience & Biobehavioral Reviews*, 35:1946–1958.

Deng, Z., Vahdat, A., Hu, H., and Mori, G. (2016). Structure inference machines: recurrent neural networks for analyzing relations in group activity recognition. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4772–4781, Las Vegas, NV. IEEE.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, 41:417–440.

Dihl, L., Testa, E., Knob, P., Lima, G., Favaretto, R. M., Alcantara, M., and Musse, S. R. (2017). Generating cultural characters based on hofstede dimensions. In *Virtual Humans and Crowds for Immersive Environments (VHCIE)*, pages 1–6, Los Angeles, CA. IEEE.

Dreyfuss, H. and Tilley, A. R. (2002). *The Measure of Man and Woman: Human Factors in Design*. John Wiley & Sons, New York, NY.

Dyaram, L. and Kamalanabhan, T. J. (2005). Unearthed: The other side of group cohesiveness. *Journal of Social Sciences*, 10:185–190.

Favaretto, R. M., Dihl, L., Barreto, R., and Musse, S. R. (2016a). Using group behaviors to detect hofstede cultural dimensions. In *International Conference on Image Processing (ICIP)*, pages 2936–2940, Phoenix, AZ. IEEE.

Favaretto, R. M., Dihl, L., and Musse, S. R. (2016b). Detecting crowd features in video sequences. In *Proceedings of Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 201–208, São José dos Campos, SP. IEEE Computer Society.

Favaretto, R. M., Dihl, L., Musse, S. R., Vilanova, F., and Costa, A. B. (2017). Using big-five personality model to detect cultural aspects in crowds. In *Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 223–229, Niteroi, RJ. IEEE.

Favaretto, R. M., Knob, P., Musse, S. R., Vilanova, F., and Costa, A. B. (2018). Detecting personality and emotion traits in crowds from video sequences. *Machine Vision and Applications*, n.d.:1–14.

Feng, L. and Bhanu, B. (2015). Understanding dynamic social grouping behaviors of pedestrians. *Journal of Selected Topics in Signal Processing*, 9:317–329.

Flötteröd, G. and Lämmel, G. (2015). Bidirectional pedestrian fundamental diagram. *Transportation Research Part B: Methodological*, 71:194–212.

Forsyth, D. (2018). *Group Dynamics*. Cengage Learning, Boston, MA.

Forsyth, D. R. (2010). *Group Dynamics*. Wadsworth, Cengage Learning, Belmont, CA.

Fridman, N., Zilka, A., and Kaminka, G. A. (2011). The impact of cultural differences on crowd dynamics in pedestrian and evacuation domains. (technical report), Bar Ilan University.

Fruin, J. J. (1971). *Pedestrian planning and design*. Metropolitan Association of Urban Designers and Environmental Planners, New Zealand, NZ.

Fukunaga, K. (1990). *Introduction to statistical pattern recognition*. Academic Press Professional, Inc., San Diego, CA.

Goldberg, L. R. (1982). *From ace to zombie: Some explorations in the language of personality*, chapter 6, pages 203–234. Lawrence Erlbaum Associates, Hillsdale, NJ.

Goldberg, L. R. (1990). An alternative description of personality: the big-five factor structure. *Journal of personality and social psychology*, 59:1216–1229.

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist*, 48:26–34.

Goldstein, S. (2015). *Cross-Cultural Explorations: Activities in Culture and Psychology*. Routledge, New York, NY.

Goode, E. (1992). *Collective behavior*. Saunders College Pub, Fort Worth, TX.

Guy, S. J., Kim, S., Lin, M. C., and Manocha, D. (2011). Simulating heterogeneous crowd behaviors using personality trait theory. In *Proceedings of the 2011 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 43–52, New York, NY. ACM.

Hall, E. T. (1966). *The Hidden Dimension*. Doubleday, Garden City, NY.

Hausdorff, F. (1962). *Set Theory*. Chelsea Publishing Company, New York, NY.

Helbing, D. (1991). A mathematical model for the behavior of pedestrians. *Behavioral Science*, 36:298–310.

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

Helbing, D. and Johansson, A. (2011). *Pedestrian, crowd and evacuation dynamics*. Springer, New York, NY.

Helbing, D., Johansson, A., and Al-Abideen, H. Z. (2007). Dynamics of crowd disasters: an empirical study. *Physical Review E*, 75:461091–461097.

Henderson, L. F. (1971). The statistics of crowd fluids. *Nature*, 229:381–383.

Hillier, B. (2002). A theory of the city as object: or, how spatial laws mediate the social construction of urban space. *Urban design international*, 7:153–179.

Hofstede, G. (2001). *Culture's consequences: comparing values, behaviors, institutions, and organizations across nations*. Sage Publications, Thousand Oaks, CA.

Hofstede, G. (2011). Dimensionalizing cultures: The hofstede model in context. *Online Readings in Psychology and Culture*, 2:3–26.

Hofstede, G., Hofstede, G., and Minkov, M. (2010). *Cultures and Organizations: Software of the Mind, Third Edition*. McGraw Hill Professional, New York, NY.

Hofstede, G., Hofstede, G. J., and Minkov, M. (1991). *Cultures and organizations: Software of the mind*. McGraw-Hill, New York, NY.

Ibrahim, M. S., Muralidharan, S., Deng, Z., Vahdat, A., and Mori, G. (2016). A hierarchical deep temporal model for group activity recognition. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1971–1980, Las Vegas, NV. IEEE.

Jacques, J. C. S., Braun, A., Soldera, J., Musse, S. R., and Jung, C. R. (2007). Understanding people motion in video sequences using voronoi diagrams. *Pattern Analysis and Applications*, 10:321–332.

Jacques Junior, J., Musse, S. R., and Jung, C. (2010). Crowd analysis using computer vision techniques. *IEEE Signal Processing Magazine*, 27:66–77.

Jelić, A., Appert-Rolland, C., Lemercier, S., and Pettré, J. (2012). Properties of pedestrians walking in line: Fundamental diagrams. *Physical Review E*, 85:36111–36120.

Jo, H., Chug, K., and Sethi, R. J. (2013). A review of physics-based methods for group and crowd analysis in computer vision. *Journal of Postdoctoral Research*, 1:4–7.

John, O. P. (1990). *The "Big Five" factor taxonomy: Dimensions of personality in the natural language and in questionnaires*, chapter 4, pages 66–100. 66–100, New York, NY.

Kapadia, M., Shoulson, A., Durupinar, F., and Badler, N. I. (2013). *Authoring Multi-actor Behaviors in Crowds with Diverse Personalities*, chapter 8, pages 147–180. Springer, New York, NY.

Kendall, D. (2016). *Sociology in our times*. Cengage Learning, Belmont, CA.

Kluckhohn, C. (1962). *Culture and behavior: collected essays*. Free Press of Glencoe, New York, NY.

Knob, P., Alcântara, M., Testa, E., Favaretto, R., Lima, G., Dihl, L., and Musse, S. R. (2018a). Generating background npcs motion and grouping behavior based on real video sequences. *Entertainment Computing*, 27:179–187.

Knob, P., de A. Araujo, V. F., Favaretto, R. M., and Musse, S. R. (2018b). Visualization of interactions in crowd simulation and video sequences. In *XVII Brazilian Symposium on Games and Digital Entertainment (SBGames)*, pages 601–610, Foz do Iguaçu, PR. SBGames.

Kocak, Y. P. and Sevgen, S. (2017). Detecting and counting people using real-time directional algorithms implemented by compute unified device architecture. *Neurocomputing*, 248:105–111.

Krüchten, C. and Schadschneider, A. (2017). Empirical study on social groups in pedestrian evacuation dynamics. *Physica A: Statistical Mechanics and its Applications*, 475:129–141.

Kuo, J. Y., Fan, G. D., and Lai, T. Y. (2016). People counting base on head and shoulder information. In *International Conference on Knowledge Engineering and Applications (ICKEA)*, pages 52–55, Singapore, CN. IEEE.

Lala, D., Thovuttikul, S., and Nishida, T. (2011). Towards a virtual environment for capturing behavior in cultural crowds. In *Sixth International Conference on Digital Information Management*, pages 310–315, Melbourn, AU. IEEE.

Li, W., Di, Z., and Allbeck, J. M. (2012). Crowd distribution and location preference. *Computer Animation and Virtual Worlds*, 23:343–351.

Lövheim, H. (2012). A new three-dimensional model for emotions and monoamine neurotransmitters. *Medical Hypotheses*, 78:341–348.

Lu, L., Chan, C., Wang, J., and Wang, W. (2017). A study of pedestrian group behaviors in crowd evacuation based on an extended floor field cellular automaton model. *Transportation Research Part C: Emerging Technologies*, 81:317–329.

Ma, Y., Li, L., Zhang, H., and Chen, T. (2017). Experimental study on small group behavior and crowd dynamics in a tall office building evacuation. *Physica A: Statistical Mechanics and its Applications*, 473:488–500.

MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297, Berkeley, CA. University of California Press.

Mahadevan, V., Li, W., Bhalodia, V., and Vasconcelos, N. (2010). Anomaly detection in crowded scenes. In *Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1975–1981, San Francisco, CA. IEEE.

Mahalanobis, P. C. (1936). On the generalised distance in statistics. In *Proceedings of the National Institute of Science of India*, pages 49–55, Odisha, IN. Calcutta.

McCrae, R. R. and Costa, P. L. (1996). *Toward a new generation of personality theories: Theoretical contexts for the five-factor model*, chapter 3, pages 51–87. Guilford Press, New York, NY.

Mccrae, R. R. and John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, 60:175–215.

Moller, M. F. (1993). A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*, 6:525–533.

Moore, S. C., Flajšlik, M., Rosin, P. L., and Marshall, D. (2008). A particle model of crowd behavior: Exploring the relationship between alcohol, crowd dynamics and violence. *Aggression and Violent Behavior*, 13:413–422.

Moussaid, M., Helbing, D., Garnier, S., Johansson, A., Combe, M., and Theraulaz, G. (2009). Experimental study of the behavioural mechanisms underlying self-organization in human crowds. *Royal Society B: Biological Sciences*, 276:2755–2762.

Murino, V., Cristani, M., Shah, S., and Savarese, S. (2017). *Group and Crowd Behavior for Computer Vision*. Elsevier Science, London, UK.

Musse, S. R. and Thalmann, D. (1997). A model of human crowd behavior : Group inter-relationship and collision detection analysis. In *Computer Animation and Simulation '97*, pages 39–51, Vienna, AT. Springer Vienna.

Narang, S., Best, A., Curtis, S., and Manocha, D. (2015). Generating pedestrian trajectories consistent with the fundamental diagram based on physiological and psychological factors. *PLOS ONE*, 10:1–17.

Nas, P. J. M. (2011). *Cities full of symbols: A theory of urban space and culture*. Amsterdam University Press, Amsterdam, NL.

Nazir, A., Enz, S., Lim, M. Y., Aylett, R., and Cawsey, A. (2009). Culture–personality based affective model. *AI & Society*, 24:281–293.

Ortony, A., Clore, G. L., and Collins, A. (1990). *The cognitive structure of emotions*. Cambridge university press, New York, NY.

Padgham, M. (2012). Human movement is both diffusive and directed. *Plos One*, 7:377–384.

Palanisamy, G. and Manikandan, T. T. (2017). Group behaviour profiling for detection of anomaly in crowd. In *International Conference on Technical Advancements in Computers and Communications (ICTACC)*, pages 11–15, Melmaurvathur, IN. IEEE.

Pan, X., Han, C. S., Dauber, K., and Law, K. H. (2007). A multi-agent based framework for the simulation of human and social behaviors during emergency evacuations. *AI & Society: Knowledge, Culture and Communication*, 22:113–132.

Pelechano, N., Allbeck, J. M., Kapadia, M., and Badler, N. I. (2016). *Simulating heterogeneous crowds with interactive behaviors*. A. K. Peters, Ltd., Natick, MA.

Predtechenskii, V. M. and Milinskii, A. I. (1978). *Planning for foot traffic flow in buildings*. Amerind Publishing, Chicago, US.

Revelle, W. and Scherer, K. R. (2009). Personality and emotion. *Oxford companion to emotion and the affective sciences*, 1:304–306.

Rodriguez, M., Sivic, J., Laptev, I., and Audibert, J. Y. (2011). Data-driven crowd analysis in videos. In *Proceedings of the International Conference on Computer Vision (ICCV)*, pages 1235–1242, Barcelona, SP. IEEE.

Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, 39:1161–1178.

Saifi, L., Boubetra, A., and Nouioua, F. (2016). An approach for emotions and behavior modeling in a crowd in the presence of rare events. *Adaptive Behavior*, 24:428–445.

Sethi, R. J. (2015). Towards defining groups and crowds in video using the atomic group actions dataset. In *International Conference on Image Processing (ICIP)*, pages 2925–2929, Quebec City, QC. IEEE.

Seyfried, A., Boltes, M., Kähler, J., Klingsch, W., Portz, A., Rupprecht, T., Schadschneider, A., Steffen, B., and Winkens, A. (2010). *Enhanced empirical data for the fundamental diagram and the flow through bottlenecks*, chapter 11, pages 145–156. Springer Berlin Heidelberg, Berlin, GE.

Seyfried, A. and Schadschneider, A. (2008). Fundamental diagram and validation of crowd models. In *Proceedings of the 8th International Conference on Cellular Automata for Reseach and Industry*, pages 563–566, Berlin, GE. Springer-Verlag.

Seyfried, A., Steffen, B., Klingsch, W., and Boltes, M. (2005). The fundamental diagram of pedestrian movement revisited. *Journal of Statistical Mechanics: Theory and Experiment*, 10:10002–10015.

Shahhoseini, Z. and Sarvi, M. (2019). Pedestrian crowd flows in shared spaces: investigating the impact of geometry based on micro and macro scale measures. *Transportation Research Part B: Methodological*, 112:57–87.

Shaikh, S. H., Saeed, K., and Chaki, N. (2016). *CAVIAR: Context Aware Vision using Image-based Active Recognition*. 2016, Kolkata, IN.

Shao, J., Kang, K., Loy, C. C., and Wang, X. (2015). Deeply learned attributes for crowded scene understanding. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4657–4666, Boston, MA. IEEE.

Shao, J., Loy, C. C., and Wang, X. (2014). Scene-independent group profiling in crowd. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2227–2234, Columbus, OH. IEEE.

Solera, F., Calderara, S., and Cucchiara, R. (2013). Structured learning for detection of social groups in crowd. In *10th IEEE International Conference on Advanced Video and Signal Based Surveillance*, pages 7–12, Krakow, PL. IEEE.

Solmaz, B., Moore, B. E., and Shah, M. (2012). Identifying behaviors in crowd scenes using stability analysis for dynamical systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34:2064–2070.

Sorokowska, A., Sorokowski, P., Hilpert, P., Cantarero, K., Frackowiak, T., Ahmadi, K., Alghraibeh, A. M., Aryeetey, R., Bertoni, A., Bettache, K., Blumen, S., Błażejewska, M., Bortolini, T., Butovskaya, M., C., F. N., Cetinkaya, H., Cunha, D., David, D., David, O. A., Dileym, F. A., Espinosa, A. C. D., Donato, S., Dronova, D., Dural, S., Fialová, J., Fisher, M., Gulbetekin, E., Akkaya, A. H., Hromatko, I., Iafrate, R., Iesyp, M., James, B., Jaranovic, J., Jiang, F., Kimamo, C. O., Kjelvik, G., Koç, F., Laar, A., Lopes, F. A., Macbeth, G., Marcano, N. M., Martinez, R., Mesko, N., Molodovskaya, N., Moradi, K., Motahari, Z., Mühlhauser, A., Natividade, J. C., Ntayi, J., Oberzaucher, E., Ojedokun, O., Omar-Fauzee, M. S. B., Onyishi, I. E., Paluszak, A., Portugal, A., Razumiejczyk, E., Realo, A., Relvas, A. P., Rivas, M., Rizwan, M., Salkičević, S., Sarmány-Schuller, I., Schmehl, S., Senyk, O., Sinding, C., Stamkou, E., Stoyanova, S., Šukolová, D., Sutresna, N., Tadinac, M., Teras, A., Ponciano, E. L. T., Tripathi, R., Tripathi, N., Tripathi, M., Uhryn, O., Yamamoto, M. E., Yoo, G., and Pierce, J. D. (2017). Preferred interpersonal distances: a global comparison. *Journal of Cross-Cultural Psychology*, 48:577–592.

Still, G. K. (2000). *Crowd dynamics*. (phd thesis), University of Warwick, Coventri, UK.

Tafjel, H. and Turner, J. C. (1986). The social identity theory of intergroup behavior. *Psychology of intergroup relations*, n.d.:7–24.

Thalmann, D. and Musse, S. (2013). *Crowd simulation*. Springer London, London, UK.

Tosato, D., Spera, M., Cristani, M., and Murino, V. (2013). Characterizing humans on riemannian manifolds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35:1972–1984.

Turner, J. C. (1981). The experimental social psychology of intergroup behavior. *Intergroup behaviour*, 66:n.d.

Van den Berg, J., Lin, M., and Manocha, D. (2008). Reciprocal velocity obstacles for real-time multi-agent navigation. In *International Conference on Robotics and Automation*, pages 1928–1935, Pasadena, CA. IEEE.

Vilanova, F., Beria., F. M., Costa, A. B., and Koller, S. H. (2017). Deindividuation: From le bon to the social identity model of deindividuation effects. *Cogent Psychology*, 4:1–21.

Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 511–518, Kauai, HI. IEEE Computer Society.

Wang, C., Zhang, H., Yang, L., Liu, S., and Cao, X. (2015). Deep people counting in extremely dense crowds. In *Proceedings of the 23rd ACM International Conference on Multimedia*, pages 1299–1302, New York, NY. ACM.

Weidmann, U. (1993). Transporttechnik der fussgänger. (technical report), Institut für Verkehrsplanung.

Weina, G., Robert, T. C., and Barry, R. (2012). Vision-based analysis of small groups in pedestrian crowds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34:1003–1016.

Wolinski, D., J. Guy, S., Olivier, A. H., Lin, M., Manocha, D., and Pettré, J. (2014). Parameter estimation and comparative evaluation of crowd simulations. *Computer Graphics Forum*, 33:303–312.

Wu, N. (2002). A new approach for modeling of fundamental diagrams. *Transportation research part A: policy and practice*, 36:867–884.

Yuan, Y., Fang, J., and Wang, Q. (2015). Online anomaly detection in crowd scenes via structure analysis. *IEEE Transactions on Cybernetics*, 45:548–561.

Zhan, B., Monekosso, D., Remagnino, P., Velastin, S. A., and Xu, L. (2008). Crowd analysis: a survey. *Machine Vision and Applications*, 19:345–357.

Zhang, J. (2012). *Pedestrian fundamental diagrams: comparative analysis of experiments in different geometries*. (ph.d. thesis), Universität Wuppertal, Jülich, DE.

Zhang, J., Klingsch, W., Schadschneider, A., and Seyfried, A. (2011). Transitions in pedestrian fundamental diagrams of straight corridors and t-junctions. *Journal of Statistical Mechanics: Theory and Experiment*, n.d.:6004–6021.

Zhang, J., Klingsch, W., Schadschneider, A., and Seyfried, A. (2012). Ordering in bidirectional pedestrian flows and its influence on the fundamental diagram. *Journal of statistical mechanics: theory and experiment*, 2012:2002–2015.

Zhang, Y., Lu, H., Zhang, L., Ruan, X., and Sakai, S. (2016). Video anomaly detection based on locality sensitive hashing filters. *Pattern Recognition*, 59:302–311.

Zhao, W., Zhang, Z., and Huang, K. (2018). Gestalt laws based tracklets analysis for human crowd understanding. *Pattern Recognition*, 75:112–127.

Zhou, B., Tang, X., Zhang, H., and Wang, X. (2014). Measuring crowd collectiveness. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36:1586–1599.

# Appendix A – VIDEO ANALYSIS DATASET AND APPLICATIONS

In addition to the geometric dimensions model proposed in this thesis, other contributions are presented in this chapter. We have built a dataset of videos with crowds of different countries. Together these videos are files with tracking, pedestrian and crowd features and geometric dimensions information. In addition, we have developed software for the detection and analysis of geometric dimensions in videos, along with a visualizer of features.

Section *A.1* presents the *Cultural Crowds* dataset, a database of videos with crowds from various countries. We show how the dataset is organized. Section *A.2* is responsible for present *GeoMind*, a software developed in Matlab App Designer with a simple interface to detect and analyze the geometrical dimensions in videos. Lastly, Section *A.3* presents a crowd features viewer.

## A.1 Cultural Crowds Dataset

Throughout the development of this work, a series of videos of crowds in different countries were selected with the intention to serve as a basis for the realization of the experiments. These videos composed a dataset called *Cultural Crowds*[1], which consists of 88 video clips with different configurations.

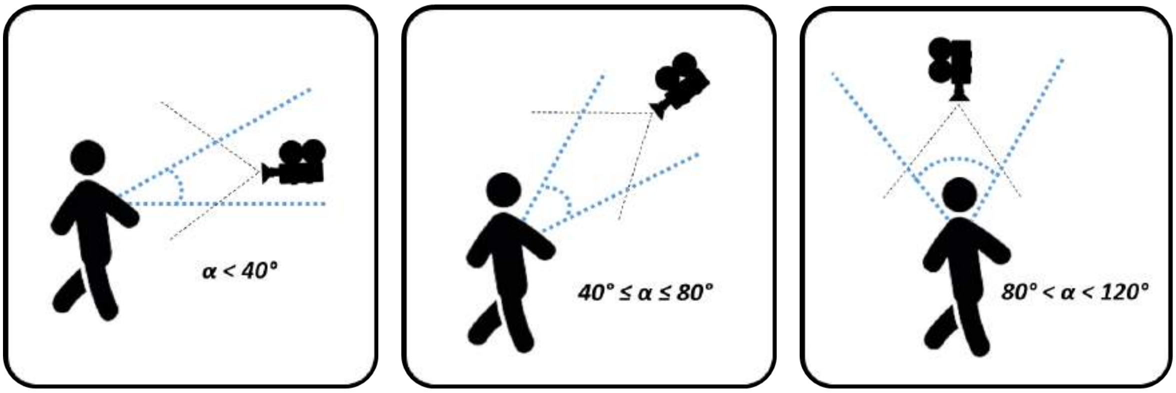
These videos, lasting between 100 and 900 frames, were collected from different public databases available on the Internet, such as Zhou et al. (2014); Shaikh et al. (2016); Rodriguez et al. (2011). The videos were cataloged in several classification criteria, such as: crowd density, camera angle, environment, nationality and type of scene.

Regarding the density aspect $\langle\Phi\rangle$ (number of pedestrians $p/m^2$), three categories were created: low, medium and high. A video falls into one of these categories based on the following:

$$Density\langle\Phi\rangle = \begin{cases} \textbf{low}, & \text{when } \langle\Phi\rangle \leq 1.5; \\ \textbf{medium}, & \text{when } 1.5 < \langle\Phi\rangle \leq 4; \\ \textbf{high}, & \text{when } \langle\Phi\rangle > 4. \end{cases} \quad (A.1)$$

[1] Cultural Crowds dataset can be accessed in http://rmfavaretto.pro.br/vhlab/datasets.php

The angle of the camera represents the angle $\alpha$ formed by the position of the camera relative to the crowd at the time the video was captured, also classified into three levels: Eye-Level, Oblique and Bird's Eye view. The characteristic of each of these levels is illustrated in Figure A.1.




(a) Eye-Level view (b) Oblique view (c) Bird's Eye view

Figure A.1: **Video classification levels**: classifications according to the angle $\alpha$ formed by the camera position in relation to the crowd: (a) Eye-Level view ($\alpha < 40°$), (b) Oblique view ($40° \leq \alpha \leq 80°$) and (c) Bird's eye view ($80° < \alpha < 120°$).

The type of the environment is related to the type of environment where the video clips were obtained. We divided the environment classification into two levels: (i) *indoor* and (ii) *outdoor*.

Nationality represents the country in which the video was recorded, in order to allow us to investigate different cultural aspects. The video clips were grouped according to the crowd scenes from where they were obtained, such as crossroad, sidewalk, running, mall, airport, stairs and escalator.

In this work, we considered only the low and medium density videos of the database, since in high density videos finding groups of people is not a trivial task. Regarding the angle of the camera, the videos captured with cameras in the frontal position were disregarded. These videos were disregarded because, when calculating the homogeneous planar projection, the error would be greater.

The videos with oblique camera angle and top were recorded with the camera positioned in high places, in these videos the view of people in the crowd is from top to bottom, suggesting a smaller error in 3D planar projection for 2D ($z$ = 0).

After these definitions of density and camera angle, 30 videos from 11 different countries were selected, distributed as follows: 3 videos from Austria - AT, 9 videos from Brazil - BR, 5 videos from China - CN , 2 videos from France - FR, 2 videos from Germany - DE, 1 video from England - UK, 3 videos from Japan - JP, 1 video from Portugal - PT, 1 video from Spain - ES, 1 video from Turkey - TR and 2 videos from the United Arab Emirates (Dubai) - AE. Figure A.2 shows the most representative frames of each one of the 30 videos used in the experiments.


Figure A.2: **Cultural Crowds dataset**: some representative frames of Cultural Crowds dataset identified by country (the two letters indicate the country and the two numbers indicate the sequence. For example, CN-05 is the 5th video from China).

Some of the videos illustrated in Figure A.2 are sequences recorded in the same environments. The videos represented in Figure A.2 are identified in the format "«xx» - «yy»", where "«xx»" indicate the two letters who represent the name of the country and "«yy»" is a sequence number that identifies a video from a specific country. To exemplify, “BR-03” is the 3rd video from Brazil and “CN-05” is the 5th video from China.

In the next section is presented a software to perform all the analysis proposed in this thesis called *GeoMind*.

## A.2 Software GeoMind

The software, called *GeoMind*[2] (abbreviated form of Geometrical Mind), was developed using Matlab App Desinger. It was design to be simple and easy to use, allowing users, with a few steps, to obtain a series of pedestrian features from video sequences, based on tracking. Fig. A.3 shows the main window of the software. It is possible to see the setup panel on the left side.

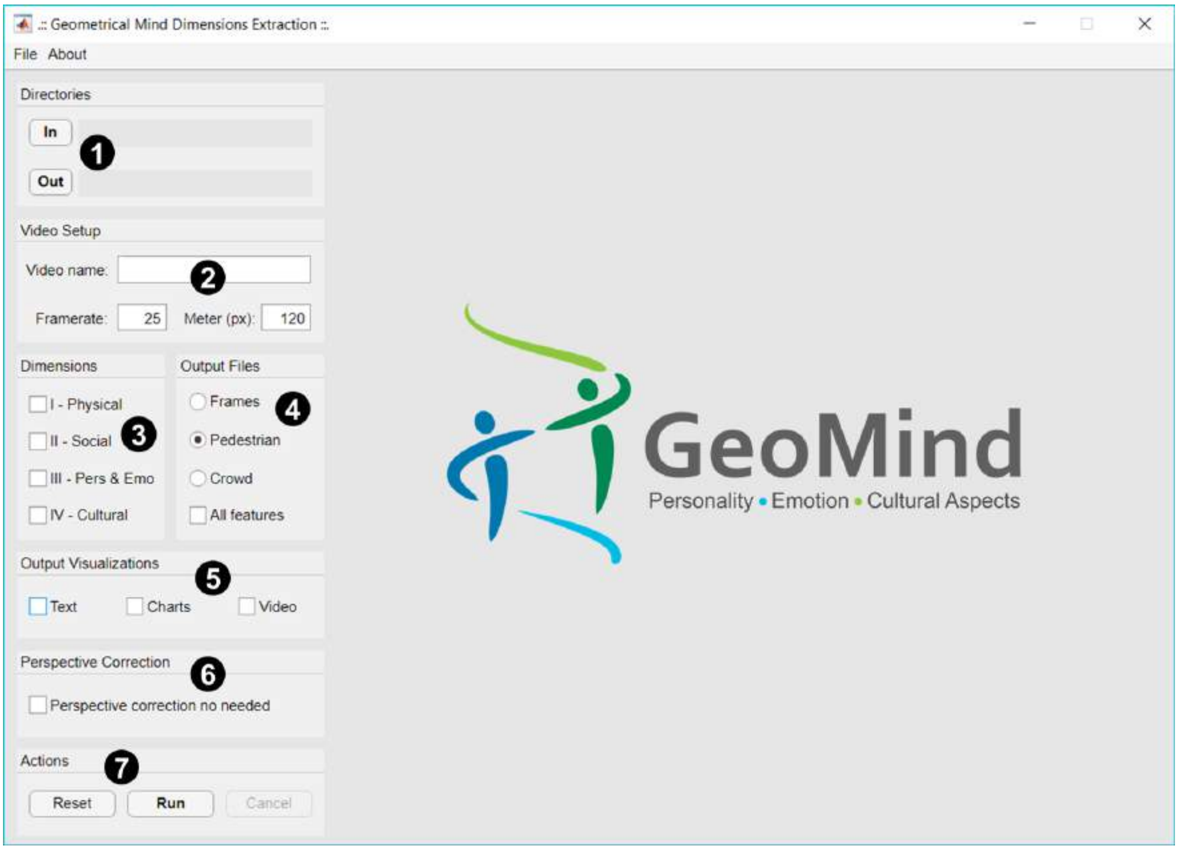



Figure A.3: **Geomind**: main window of *GeoMind*.

[2]Download and more information about how to use *GeoMind* can be found at http://rmfavaretto.pro.br/geomind.

The setup panel (left side of window in Fig. A.3) contains the input and output configurations of the video processing:

**1. Directories**: the input directory must have the video frames (pictures) and the tracking file. The output files generated by *GeoMind* will be stored in the output directory;

**2. Video Setup**: users must specify a video name, its framerate and how many pixels represent one meter in the video;

**3. Dimensions**: the features were grouped into four dimensions, accordingly to proposed in this thesis: *I - Physical* with geometrical features, such as speed and angular variation; *II - Social* with features related to groups and social behaviours, as collectivity and socialization; *III - Personal and Emotional* with regard to OCEAN personality and OCC emotion models; and *IV - Cultural* with respect to Hofstede cultural Dimensions. The users have to select at least one dimension to be computed and saved in the output directory;

**4. Output Files**: in this section, users have to select the frequency in which the features are saved in the output files. In addition, they can check the option “All features”, outputting a file with all available information about pedestrians;

**5. Output Visualizations**: users can choose how they want to generate the information about pedestrians, as text in *.txt* files, as plot in charts or in a video;

**6. Perspective Correction**: optionally, users can use a version of the tracking with perspective correction, reducing errors during calculations of pedestrian positions in the video. If this correction is not necessary, just check the “Perspective correction no needed” option;

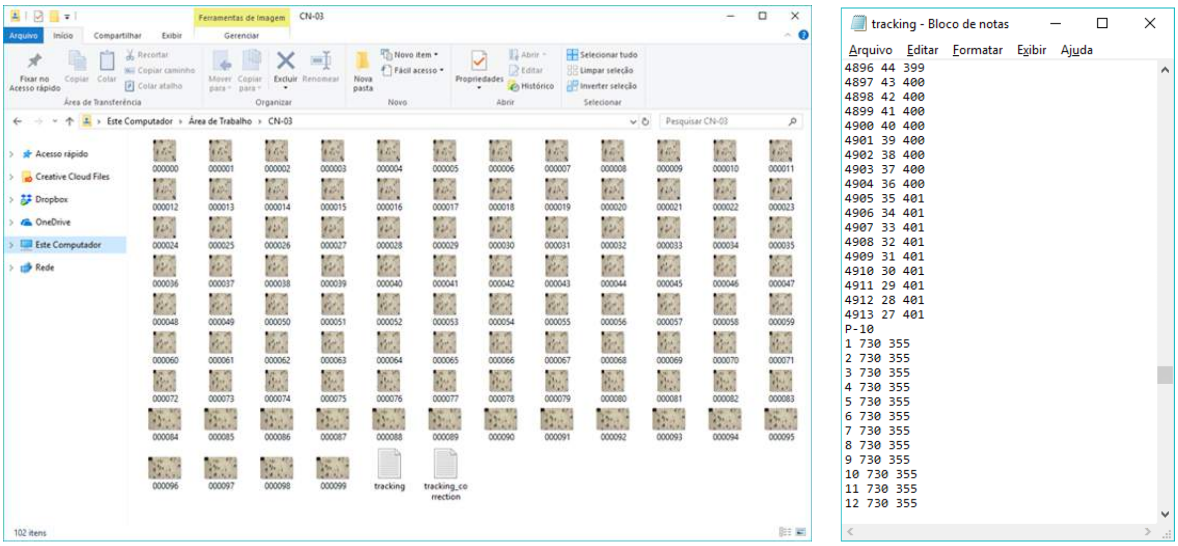
**7. Action Buttons**: once the user filled all the fields in the setup panel, he has 3 options to choose: *Run* to start the video processing; *Reset* to restore default values for each predefined form field; and *Cancel*, to stop the processing in progress.

Input Files Format

The input directory must have the video frames and the tracking file. Each frame of the video sequence must be extracted as an *.jpg* picture. The frame names must have a sequence of six digits, starting by *000000.jpg* (for example, the frame names from a video with 853 frames should be *000000.jpg*, . . . , *000852.jpg*). The tracking file (named as *tracking.txt*) must have the 2D positions of each pedestrian at each frame.

The tracking file keeps the positions of each pedestrian, in the following format: pedestrian identification P-«id», where «id» is a sequential number starting in zero; followed by the tuple $<$F X Y$>$, indicating the frame $F$ and the position ($XY$) of that pedestrian in that frame. An example of the tracking file can be found in Figure A.4(b). In addition, a tracking file can be used with perspective correction. In this case, the *tracking_correction.txt* file must have the same format as *tracking.txt*.

Figure A.4 illustrates the input files needed by *GeoMind*. Figure A.4(a) shows the input directory, where it is possible to see 100 frames of the video (files *000000.jpg*, . . . , *000099.jpg*) and the tracking files (*tracking.txt*, with the positions in image coordinates and *tracking_correction.txt*, with positions in world coordinates, after a perspective correction). Figure A.4(b) shows an example of a tracking file. It is possible to see the last positions of pedestrian *P-9* and the first positions of pedestrian *P-10*.




(a) Input files (b) Tracking file

Figure A.4: **Examples of *GeoMind* input files**: input directory with the tracking files and frames of the video (a) and an example of *tracking.txt* file (b).

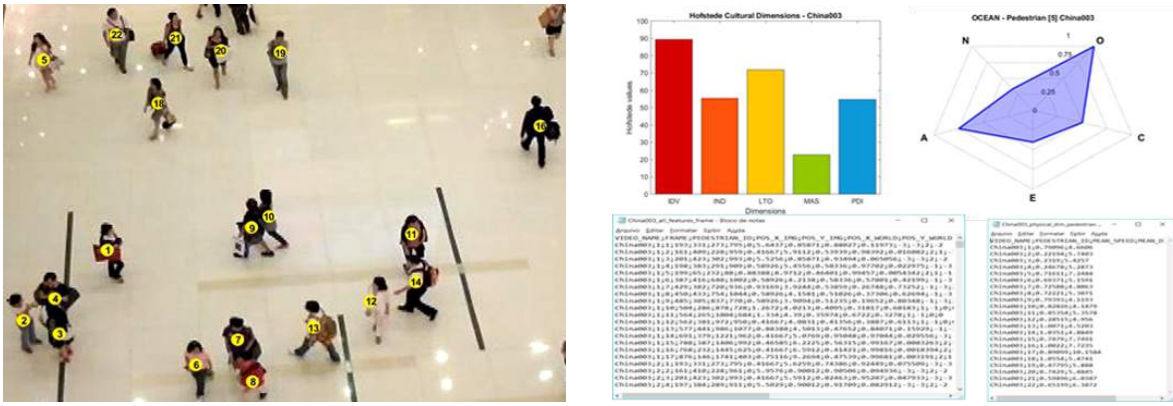
In the next section we present the results and the files generated by *GeoMind* after the video processing.

Fig. A.5 shows a video summary after the software processing. This summary is divided in five areas. Area **a** shows the video information, such as number of frames, number of pedestrians and number of groups. In this area any individual can be select to see its features (“Pedestrian 1" is selected).


Figure A.5: **Geomind**: video summary after video processing in *GeoMind*.

Area **b** shows an image containing the summary on the *Physical* features, presenting the mean distance of selected pedestrian to the others, mean speed, an indication if the pedestrian is part of a group or not and a plot of its speed over time. In area **c** thee are the features related to the *Social* dimension. It shows data about Isolation and socialization levels and a plot on the collectivity of the selected pedestrian over time. This section also brings information about groups found in the video, such as: number of grouped and ungrouped pedestrians, mean size of the groups (number of pedestrians), mean cohesion, mean area and mean distance between the pedestrians in the group.

Area **d** shows the *Personal and Emotion* dimensions features, plotting the emotion values and personalities (Big-Five) of the selected pedestrian in the interval [0, 1]. Finally, area **e** is responsible for the *Cultural* features, plotting the Hofstede Cultural Dimensions of the video. Fig. A.6 shows some examples of files generated by *GeoMind*: a frame of a video with the pedestrians' ID in Fig. A.6(a) and examples of charts and text files with pedestrians' features in Fig. A.6(b).




(a) Pedestrians' ID

(b) Plots and text files

Figure A.6: **Examples of files generated by *GeoMind***: video with pedestrians ID (a) and plots and text files with features (b).

One of the files generated by *GeoMind* (*«video»_all_features_frame.txt*, where «video» indicate the name given to the video in the setup panel at proccessing time) can be used as an input in a visualizer tool. That visualizer tool is presented in next section.

### A.3 Visualizing Cultural Aspects

As mentioned before, one of the outputs generated by the *GeoMind* software is the *all_features_frame.txt* file from a specific video. This file contains all the information about each of the pedestrians present in the video at each frame. This file serves as input to a viewer, which allows, in a simulation environment, a way of visualizing a series of features, such as: cultural aspects, personality traits, social aspects, emotions, socialization, isolation, collectivity, among other characteristics.

This application[3] was developed using the Unity3D[4] engine, with *C#* programming language. The viewer allows the users to rewind, accelerate and stop the simulated video through a time controller, so that the user can observe something that he finds interesting several times, at any time. Figure A.7 shows the main window of the viewer.

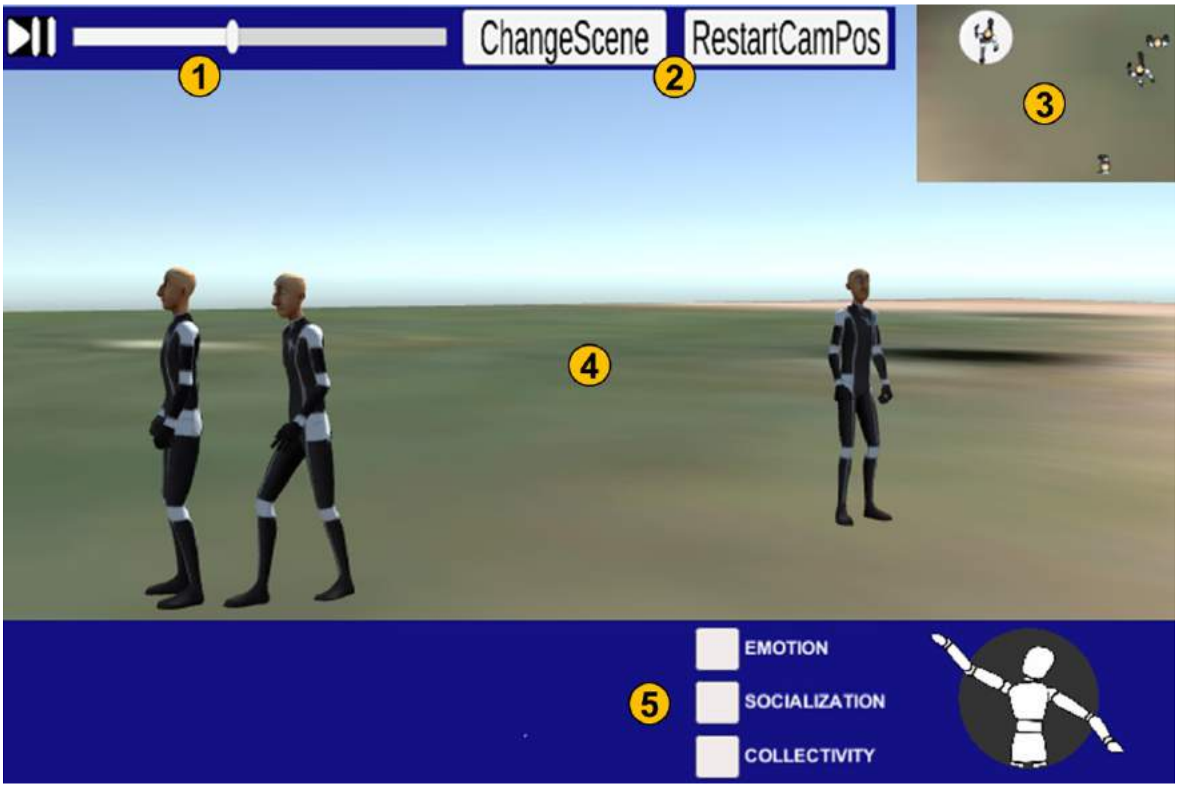



Figure A.7: **Main window of the viewer**: visualization of a file output in *GeoMind* from a determined video.

This environment has three modes of visualization: (i) first-person visualization, where the user can visualize what is happening in the environment from the perspective of a pedestrian present in the video; (ii) top view, where it is possible to observe the movement of the pedestrians through a perspective from above of the video environment; and (iii) an oblique view, where the user can see the entire environment. The viewer is divided in five parts, as follows:

1. **Time controller**: in the area 1, it is possible to see the button with the start, stop and continue simulation playback functions, together with the frame control bar;

[3] I would like to thank Victor Flavio Araujo, master student and colleague at VHLAB who developed this viewer during his research.

[4] Unity3D is available at https://unity3d.com/

**2. Scene setup**: in area 2 there are the *ChangeScene* and *RestartCamPos* buttons, respectively, to go to the viewer's home screen (where the user can load the data file of another video) and restart the camera position for viewing in first person, if it is fixed to a agent, to an initial default position;

**3. Top-view cam**: area 3 shows the top view of the environment, as if the user were looking at the crowd through the top;

**4. First-person cam**: area 4 shows first-person view from the viewpoint of a previously selected agent. This agent is highlighted in area 3;

**5. Features panel**: area 5 is responsible for the features panel, where the users can see up to four selected agents and their features. In addition, it is possible to activate the visualization of the data related to the emotion, socialization and collectivity of agents.

The visualization of agent characteristics is shown in the features panel, illustrated in Figure A.7, area 5. This panel is hidden, only visible on the screen if the mouse cursor passes through the lower region of the screen. In this panel, there are three check-boxes: (i) emotion, (ii) socialization, and (iii) collectivity. The function of these boxes is to enable and disable the visualization of these characteristics in the agents. Figure A.8 shows all possible icons that are related to the three options.

This view consists of a set of icons that are displayed at the top of the agents in the simulation. For example, by the time the user selects the emotion field, icons representing the emotions (anger, fear, happiness and sadness) of each agent are displayed on the top of each agent. This icons are displayed in Figure A.8(g-j).

Figure A.9 shows an example of a video loaded in the viewer. The viewer allows the selection of up to four agents, which are present in the current frame, by right clicking on the humanoids that the user wishes to select. For each selected agent, the color of his clothes is changed and his information is fixed in the features panel, represented by the same color of the clothes and by an identifier (for example *Agent10*, who is highlighted in green).

Besides the agent identifier are the representative icons of their characteristics: speed, whether the agent is walking or running in the current frame; collectivity, whether the agent is collective or not; socialization, whether the agent is sociable or isolated; and emotion, whether the agent is angry, happy, sad, or afraid. As an example, *Agent10* (highlighted in green in Figure A.9) which is running, is

(a) Walking (b) Running (c) Socialization (d) Isolation (e) Individualism

(f) Collectivism (g) Anger (h) Fear (i) Happiness (j) Sadness

Figure A.8: **Icons from the features of the viewer**: possible icons shown in the features panel from a determined agent. In (a) and (b) are illustrated the icons that represent the speed of the agent in each frame, respectively, walking and running. In (c) and (d) are illustrated the icons that represent whether the agent is sociable or isolated in that frame. In (e) and (f) are illustrated the icons that represent if the agent is collectivist or individualistic. From (g) to (f) are illustrated the icons that indicate the emotions (anger, fear, happiness, and sadness) of the agent in the current frame.

not a collective agent, is isolated and happy. All possible icons are presented in Figure A.8.

Besides that, the viewer also provides a radial menu to show the features' values of a selected pedestrian. For this, it is enough that the user clicks on the identifier of a certain agent in the panel of features and the radial menu will appear. Figure A.10 shows an example of the menu.

The features illustrated in Figure A.10 were presented in seven categories: *I - Speed*, *II - Collectivity*, *III - Interpersonal Distance*, *IV - Socialization and Isolation*, *V - Hofstede Cultural Dimensions*, *VI - Big-Five personalities* and *VII - Emotions*. in that example, the OCEAN personalities of *Agent6* (highlighted in red) were represented in a graphical way, considering the max value of each dimension.

In addition to viewing spontaneous videos, the viewer also have a module which accepts videos from controlled experiments, such as the Fundamental diagram experiment. Figure A.11 shows the visualization from a video of the FD experiment, recorded in Brazil with 15 agents. In Figure A.11(a), the oblique view is shown, where the user has a more general view of the experiment. In Figure A.11(b), the first person view is shown, where the user can feel part of the experiment, with


Figure A.9: **Emotion analysis in the viewer**: an example of the emotions shown in the top of each agent. In addition, four agents were select (highlighted with different colors), where is possible to see its features in the panel.

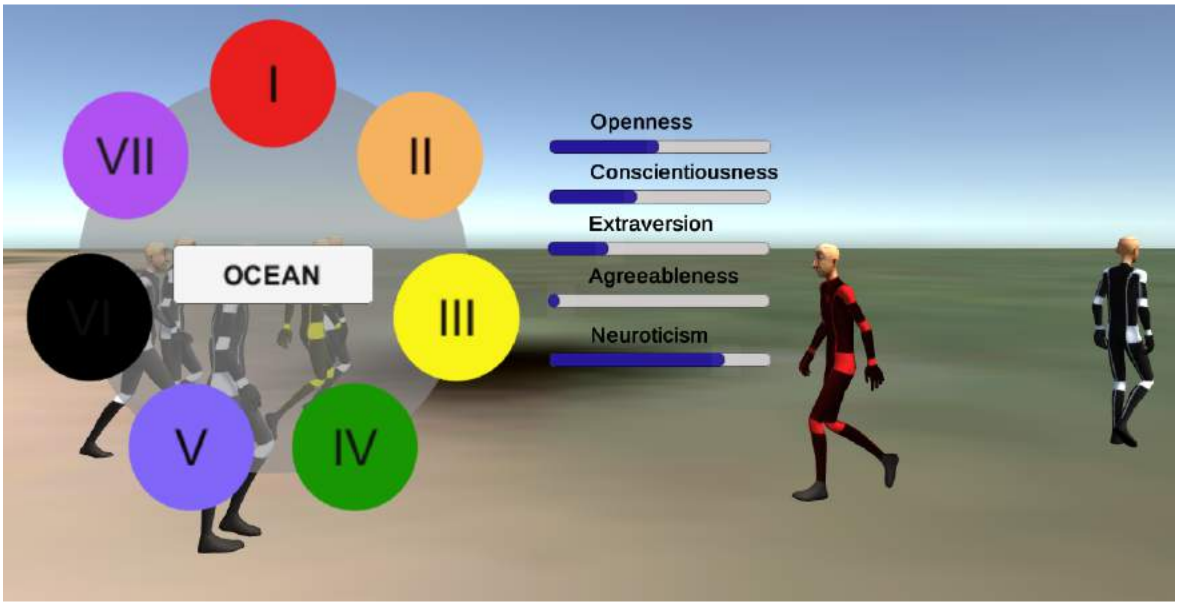
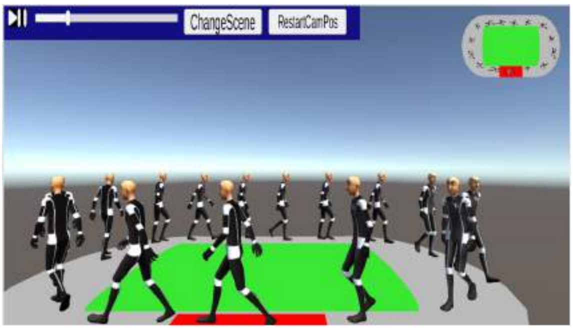
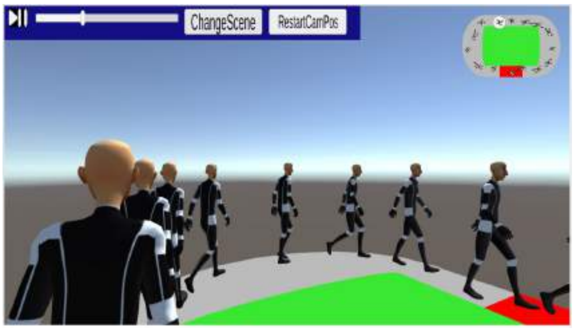



Figure A.10: **Radial menu of features**: an example of the personalities shown in the radial menu from a selected pedestrian.

the view of one of the agents (highlighted in the top view area). In both cases, the user can see a top view of the experiment in the upper right corner of the figures.




(a) Oblique view




(b) First-person view

Figure A.11: **Visualization from a video of the FD experiment**: different angle of visualizations: (a) oblique view and (b) first-person view.

## A.4 Appendix Remarks

In addition to the *Big-Four Geometrical Dimensions* presented in previous chapters, in this appendix we presented other contributions of this thesis. Firstly we presented a dataset of videos of crowds from several countries, referenced as *Cultural Crowds Dataset*. The dataset was structured under several aspects, such as crowd density (*low*, *medium* and *high*), camera position in relation to the crowd (*eye-level*, *oblique* and *bird's eye*), type of environment (*indoor* and *outdoor*), scene (*crossroad*, *sidewalk*, *mall*, among others) and the video nationality. All the videos in the dataset are accompanied by tracking and feature files.

In the sequence, we presented our software called *GeoMind*, which serves as a tool to detect and analyze geometric dimensions in video sequences, but mainly to be a proof-of-concept of our model. *GeoMind* was built in Matlab App Designer and presents a intuitive interface. Based on the frames of the video and the tracking files, the software generates a series of analysis regarding the geometrical dimensions proposed in this thesis.

Finally, we present the development of a crowd-cultural data viewer[5], which allows the users to analyze and extract pedestrian behavior information. With three modes of visualizations (*first-person*, *oblique* and *top view*), the viewer can be use as a complement to GeoMind software.

[5]As we stated before, this viewer was developed by the collaborator Victor Araujo, a master student at VHLab and not by the author of this thesis.

# Appendix B – PUBLICATIONS

This appendix presents the relation of publications obtained during the development of this research. Section *B.1* shows a list of already published researches, including conference and journal papers. Section *B.2* shows a list of submitted papers to conferences and journals without a decision by the moment of delivery of this manuscript. Section *B.3* describes a book we are finalizing the production with Spring Nature.

## B.1 Published Research

**Detecting personality and emotion traits in crowds from video sequences**.
***Rodolfo Migon Favaretto**, Paulo Knob, Soraia Raupp Musse, Felipe Vilanova and Ângelo Brandelli Costa.* Journal Machine Vision and Applications (Springer). Pages 1–14. 2018.
DOI: https://doi.org/10.1007/s00138-018-0979-y.

**Generating background NPCs motion and grouping behavior based on real video sequences**.
*Paulo Knob, Marlon Alcântara, Estêvão Testa, **Rodolfo Favaretto**, Gabriel Lima, Leandro Dihl and Soraia Raupp Musse.* Journal Entertainment Computing (Elsevier). Volume 27, pages 179–187. 2018a.
DOI: https://doi.org/10.1016/j.entcom.2018.06.003.

**Visualization of Interactions in Crowd Simulation and Video Sequences**.
*Paulo Knob, Victor Flavio de Andrade Araujo, **Rodolfo Migon Favaretto** and Soraia Raupp Musse.* Proceedings of XVII Brazilian Symposium on Games and Digital Entertainment (SBGames), Foz do Iguaçu, PR. 2018b.
LINK: http://www.sbgames.org/sbgames2018/proceedings-eng.

**Using Big Five Personality Model to Detect Cultural Aspects in Crowds**.
***Rodolfo Migon Favaretto**, Leandro Dihl, Soraia Raupp Musse, Felipe Vilanova and Angelo Brandelli Costa.* Proceedings of 30th Conference on Graphics, Patterns and Images (SIBGRAPI), Niteroi, RJ. 2017.
DOI: https://doi.org/10.1109/SIBGRAPI.2017.36.

**Generating Cultural Characters based on Hofstede Dimensions**.
*Leandro Dihl, Estêvão S. Testa, Paulo Knob, Gabriel L. B. da Silva, **Rodolfo M. Favaretto**, Marlon F. de Alcântara and Soraia R. Musse.* Proceedings of Virtual Humans and Crowds for Immersive Environments (VHCIE), Workshop of VR, Los Angeles, CA. 2017.
DOI: https://doi.org/10.1109/VHCIE.2017.7935621.

**Using group behaviors to detect Hofstede Cultural Dimensions**.
***Rodolfo M. Favaretto**, Leandro Dihl, Rodrigo Barreto and Soraia Raupp Musse.* Proceedings of IEEE International Conference on Image Processing (ICIP), Phoenix, AZ. 2016a.
DOI: https://doi.org/10.1109/ICIP.2016.7532897.

**Detecting Crowd Features in Video Sequences**.
***Rodolfo Migon Favaretto**, Leandro Lorenzett Dihl and Soraia Raupp Musse.* Proceedings of 29th Conference on Graphics, Patterns and Images (SIBGRAPI), São Paulo, SP. 2016b.
DOI: https://doi.org/10.1109/SIBGRAPI.2016.036.

**Generating Background Population for Games based on Real Video Sequences**.
*Marlon Alcantara, Estevao Testa, Gabriel Lima da Silva, **Rodolfo Favaretto**, Leandro Dihl and Soraia Musse.* Proceedings of XV Brazilian Symposium on Games and Digital Entertainment (SBGames), São Paulo, SP. 2016.
LINK: http://www.sbgames.org/sbgames2016/page/anais.

## B.2 Ongoing publications

**Video Analysis and Simulation of Cultural Behaviours in Video Sequences**.
***Rodolfo Migon Favaretto**, Soraia Raupp Musse, Paulo Knob, Estêvão Testa, Leandro Dihl, Felipe Vilanova and Angelo Brandelli Costa.* The Visual Computer (TVCJ). (Journal Elsevier). 2019
Status: **SUBMITTED**.

**GeoMind: a Video Analysis Application to Detect Personality, Emotion and Cultural Aspects of Pedestrians**.
***Rodolfo Migon Favaretto**, Soraia Raupp Musse, Felipe Vilanova and Angelo Brandelli Costa.* IEEE International Conference on Image Processing (ICIP), Taipei - Taiwan. 2019.
Status: **SUBMITTED**.

**Investigating Cultural Aspects in the Fundamental Diagram using Convolutional Neural Networks and Virtual Agents Simulation**.
***Rodolfo Migon Favaretto**, Soraia Raupp Musse, Felipe Vilanova and Angelo Brandelli Costa.* International Conference on Computer Animation and Social Agents (CASA), Paris - France. 2019.
Status: **SUBMITTED**.

**How much do you perceive this? An analysis on perceptions of geometric features, personalities and emotions in virtual humans**.
*Victor Flavio Araujo, **Rodolfo Migon Favaretto** and Soraia Raupp Musse.* Intelligent Virtual Agents (IVA), Paris - France. 2019.
Status: **SUBMITTED**.

**Big-Four Geometrical Dimensions: an Approach to Detect Personality, Emotion and Cultural Aspects in pedestrians from Crowds**.
***Rodolfo Migon Favaretto**, Soraia Raupp Musse, Felipe Vilanova and Angelo Brandelli Costa.* Frontiers in Psychology. Journal. 2019.
Status: **DRAFTING**.

### B.3 Springer Nature Book

The content of this thesis is part of a book entitled "**Emotion, Personality and Cultural Aspects in Crowds: Towards a Geometrical Mind**". The book is being developed in partnership with professors Soraia Raupp Musse (Computer Science department) and Angelo Brandelli Costa (Psychology department). We are currently finalizing the writing of the text. We signed a contract with Springer Nature[1] and the book will be delivered in the middle of April.

[1] https://www.springernature.com/gp/products/books

Pontifícia Universidade Católica do Rio Grande do Sul
Pró-Reitoria de Graduação
Av. Ipiranga, 6681 - Prédio 1 - 3º. andar
Porto Alegre - RS - Brasil
Fone: (51) 3320-3500 - Fax: (51) 3339-1564
E-mail: prograd@pucrs.br
Site: www.pucrs.br