PUCRS

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

VICTOR FLÁVIO DE ANDRADE ARAUJO

STUDY AND EVALUATION OF HUMAN PERCEPTION ON VIRTUAL HUMANS AND CROWDS

Porto Alegre

2020

PÓS-GRADUAÇÃO - *STRICTO SENSU*

Pontifícia Universidade Católica
do Rio Grande do Sul

PONTIFICAL CATHOLIC UNIVERSITY OF RIO GRANDE DO SUL
SCHOOL OF TECHNOLOGY
GRADUATE PROGRAM IN COMPUTER SCIENCE

STUDY AND EVALUATION OF HUMAN PERCEPTION ON VIRTUAL HUMANS AND CROWDS

VICTOR FLÁVIO DE ANDRADE ARAUJO

Dissertation presented as partial requirement for obtaining the degree of Master in Computer Science at Pontifical Catholic University of Rio Grande do Sul.

Advisor: Prof. Soraia Raupp Musse

Porto Alegre
2020

Ficha Catalográfica

A553s Andrade Araujo, Victor Flávio de

Study and evaluation of human perception on virtual humans and crowds / Victor Flávio de Andrade Araujo . – 2020.
97.
Dissertação (Mestrado) – Programa de Pós-Graduação em Ciência da Computação, PUCRS.

Orientadora: Profa. Dra. Soraia Raupp Musse.

1. Human Perception. 2. Crowds. 3. Interactions. 4. Cultural Features. 5. CG Characters. I. Raupp Musse, Soraia. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da PUCRS
com os dados fornecidos pelo(a) autor(a).
Bibliotecária responsável: Clarissa Jesinska Selbach CRB-10/2051

Victor Flávio de Andrade Araujo

Study and evaluation of human perception on virtual humans and crowds

This Dissertation has been submitted in partial fulfillment of the requirements for the degree of Master 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 4, 2020.

COMMITTEE MEMBERS:

Prof. Dr. Bruno Feijó (PUC-Rio)

Profa. Dra. Isabel Harb Manssour (PUCRS)

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

ESTUDO E AVALIAÇÃO DA PERCEPÇÃO HUMANA EM HUMANOS VIRTUAIS E MULTIDÕES

RESUMO

O comportamento do ser humano depende de sua percepção do mundo. Comportamentos tem sido estudados por muito tempo, em diferentes aspectos, em particular, pesquisa em Processamento Gráfico e Visual. Além de estudados, comportamentos podem ser simulados, bem como analisados em vídeos, e em ambos casos estudos de comportamentos reais são de grande valia. A área de Simulação de Multidões encontra-se neste contexto, onde o objetivo é simular e prover comportamentos realistas. Um estudo correlato é a área de percepção humana a respeito de humanos virtuais, grupos e multidões. A percepção humana é essencial para identificar se uma outra pessoa está chorando, ou está com raiva, ou qualquer outra característica no que tange a forma, movimentação ou aspectos culturais dos humanos virtuais simulados. Com todas as pesquisas sobre comportamento humano e com a evolução da Computação Gráfica, a indústria de entretenimento tem criado personagens e rostos muito realistas, as vezes causando estranheza e sentimentos involuntários na percepção humana. Nesse sentido, este trabalho tem como objetivos realizar estudos e avaliações das percepções humanas sobre humanos e multidões virtuais. Nós dividimos essas análises em três etapas: *i)* Percepção de interações de pessoas em multidões; *ii)* Percepção de características geométricas e culturais em multidões; *iii)* Percepção sobre personagens de CG. Com relação a primeira etapa, simulamos interações entre agentes virtuais através de suas percepções espaciais, identificamos interações (usando os mesmos parâmetros de interações dos agentes virtuais nas simulações) entre pessoas em sequências de vídeos reais, e fizemos visualizações interativas para facilitar as análises dos dados gerados pelas interações (tanto das simulações quanto dos vídeos). Na segunda etapa, analisamos a percepção humana através de questionário sobre dados extraídos geometricamente (distância entre pessoas, velocidade das pessoas, variação angular e densidade), além de dados não geométricos, como personalidades e emoções. Na terceira etapa, analisamos a percepção humana sobre personagens criados com CG (filmes, jogos, animações, e etc) a fim de responder questões levantadas pelos efeitos do *Uncanny Valley* (Vale da Estranheza), também através de questionários.

Palavras Chave: Percepção Humana, Multidões, Interações, Características Culturais, Personagens de CG.

STUDY AND EVALUATION OF HUMAN PERCEPTION ON VIRTUAL HUMANS AND CROWDS

ABSTRACT

The behavior of the human being depends on his perception of the world. Behaviors have been studied for a long time, in different aspects, in particular research in *Graphic* and *Visual Processing*. In addition to being studied, behaviors can be simulated, as well as analyzed in videos, and in both cases studies of real behaviors are of great value. The *Crowd Simulation* area is in this context, where the objective is to simulate and provide realistic behaviors. A correlate study is the area of human perception regarding virtual humans, groups and crowds. Human perception is essential to identify whether another person is crying, or is angry, or any other characteristic regarding the shape, movement or cultural aspects of the simulated virtual humans. With all the research on human behavior and the evolution of *Computer Graphics (CG)*, the entertainment industry has created very realistic characters and faces, sometimes causing strangeness and involuntary feelings in human perception. In this sense, this work aims to carry out studies and evaluations of human perceptions about human beings and virtual crowds. We divided these analyzes into three parts: *i)* Perception of interactions in crowds, *ii)* Perception of geometric and cultural features in crowds, *iii)* and Perception of *CG* characters. Regarding the first part, we simulate interactions between virtual agents through their spatial perceptions, identify interactions (using the same parameters of interactions of virtual agents in the simulations) between people in real video sequences, and made interactive visualizations to facilitate the analysis of the data generated by interactions (both, in simulations and videos). In the second part, we analyze human perception through a questionnaire about geometrically extracted data (distance between people, people's speed, angular variation and density), in addition to non-geometric data, such as personalities and emotions. In the third part, we analyze the human perception of characters created with *CG* (movies, games, animations, etc.) in order to answer questions raised by the effects of *Uncanny Valley*, also through questionnaires.

Keywords: Human Perception, Crowds, Interactions, Cultural Features, CG Characters.

LIST OF FIGURES

CONTENTS

1. INTRODUCTION

The study of human behavior is a subject of great scientific interest and probably an inexhaustible source of research. In particular, crowd behavior has been investigated in the context of several applications and a variety of purposes. In the context of crowd analysis, Jacques et al.[36] discussed many methods to extract data about crowds in video sequences. Many are the applications, but certainly one of the most relevant concerns the safety of pedestrians in complex buildings or in mass events. Many methodologies to detect groups and crowd events have been proposed in literature and achieved results showing that groups, social behaviors and navigation aspects can be successfully detected in video sequences. For example, counting people in crowds [8], study of social groups [9], understanding of group and crowd behaviors [58], among others. Other authors [48, 41] have recently presented topics in crowd analysis that aim to find out patterns in people behavior. Although the applications can be different, in general, the methods are focused on finding relevant data in crowd motion and behavior in video sequences, in automatic or semi-automatic ways.

In the context of simulation, crowd behavior has also been a highly explored topic of research. It is used to simulate the movement of several virtual agents in an environment area, like a square or an amusement park. It can also be used to simulate the flow of the crowd in complex environments, like people leaving a soccer stadium after a match. Also, crowd models can be used for urban planning, determining the level of comfort of each agent in a public space. One of the models developed for crowd simulation is *BioCrowds* [7], which was inspired on the computational model of leaf growth [53], where the structures known as "auxins" guide the leaf node rib growth and hence the leaf itself. As the spatial distribution of the markers (auxins) indicates where an agent can move, each agent respects the personal space of the others, creating a simulation without collisions, capable of dealing with the presence of other agents and obstacles.

If, on the one hand, researchers are interested on extracting data in video sequences, on the other hand, researchers are interested in replicating behaviors in simulations with a considerable level of realism. However, a still incipient area in this context seems to be the investigation of human perception in scenes of groups and crowds [59]. It includes people's perception regarding geometric information, but also other kind of information as personality traits, emotional status and etc. Some questions are still opened, for instance: Should the qualitative analysis agree with automatic methods? Are measurable errors among what humans observe and the quantitative analysis? This is important because many qualitative assessments are made to validate or support the computational analysis (or simulation). The research question of this work is: what can we learn about human perception of visualizing data obtained in video footage or simulating groups and crowds?

In order to address this aspect, we proceed with three topics of investigations. Firstly, we investigated some known visualization methods to present data about simulations and video sequences. In this case, we focused on geometric information (positions) that change as people interacts in tested scenarios. Our goal is to study the interaction phenomenon which occurs between people, because it is a common event that includes casual conversation, people crossing each other, among others possibilities. For this, we developed some techniques and obtained results as described in Section 3.

The second investigated topic includes visualization of people information that can not be explicitly viewed in video sequences as personalities traits and emotions. In literature, there are some work involving the analysis of cultural features, such as analysis of the impact of groups on crowds through human perceptions [59], simulation of crowds through behaviors based on personality and emotions traits [16], visualization of personality traits through social media [29], visualization and understanding of personal emotional style [63], visualization of personal records [52], among others. Recently, studies have used geometric features to analyze cultural aspects in crowds. Favaretto et al. [20] used group behaviors to detect cultural aspects according to Hofstede [34]. In other investigations, Favaretto et al. investigated cultural aspects using controlled experiment videos (related to Fundamental Diagram [10]) and spontaneous videos from various countries, using geometrical features [22], Big-Five personality [21] and OCC emotion [24] models. Section 4 discusses our contribution to visualize cultural, emotional and personality data of groups and individuals. In addition, some obtained results are discussed when analysing human perception.

As previously discussed, human perception has been very important in *Computer Graphics* (*CG*) and, in addition to helping to study human behavior, it and can be very important in the evolution of virtual humans and realistic faces [37, 46]. Increasingly, real actors have been replaced by *CG* characters and often this substitution is not even perceived by the public, but sometimes there are still some perceived oddities, such as the movement of the mouth and eyes [51]. According to Mori [49], robots made to appear too similar to real humans can fall into the "*Uncanny Valley*", where a too high degree of human realism evokes an eerie feeling in the viewer. Indeed, the effect of the *Uncanny Valley* hypothesis on *CG* characters has become increasingly influential in scientific studies [40, 39, 64, 50, 56]. So, the third aspect related to perception studied in this work is the effect of *Uncanny Valley* in *CG* characters, as presented in Section 5. In this case, we evaluated faces in images, videos and characters in interactive environments, containing individuals and groups.

1.1 Objectives

The main objective of this work is to study and evaluate human perception in *CG* scenarios containing virtual humans. To do so, we have considered three specific topics: *i)* To study and evaluate people interactions, in video sequences and simulations, using visualization techniques; *ii)* to evaluate people's perception of crowds properties (geometric and non-geometric data) and *iii)* to study people's perception of characters with *CG*. The following subsections specify these topics:

1.1.1 Perception of People Interactions in Crowds

As already mentioned, interactions are recurrent phenomena among people. We visualize interactions among agents in crowds simulated by *BioCrowds* [7], and among people in real videos from the *Cultural Crowds* data set[1] [20]. This process was divided into three steps: *i)* we simulate interactions among agents; *ii)* we used data from video sequences with people walking and used a similar method as simulations to identify interactions among them; and *iii)* we apply some existing data visualization methods to present the generated output, both for simulations and for video sequences. Using some visualization techniques, it is possible to better understand information regarding motion and interaction of agents/individuals in the scenes. This content can be seen in Section 3. While the viewer facilitates the process of understanding location of interactions and people motion, parameters like internal parameters such as personalities and emotion are not visible. So, we proceeded with the development of a tool to provide animated visualization of individuals and group parameters.

1.1.2 Perception of Cultural Features in Crowds

We investigate how people perceive the geometric features (for example, density data, distances and velocities) and non-geometric features (for example, cultural features such as personality traits and emotions), computed with the data extracted of pedestrians from videos of crowds. For this, we use the videos of *Cultural Crowds* data set, which contains videos of crowds from different countries, with pedestrians walking in different scenarios. Therefore, the data set contains the tracking files with the pedestrian positions and provides also personality and emotion information of those pedestrians, which was obtained using the *GeoMind Model* [25]. For the performed experiments, we use the tracked positions in a simulated environment where agents were visualized as identical virtual humans. The goal is to focus on their behavior and not being distracted by other visual features. In our analysis, the participants were asked to answer questions to identify if they can perceive geometric features as distances/speeds as well as emotions and personalities in video sequences, when pedestrians are represented by virtual humans. In particularly, and very important to this analysis, is to understand that our focus is on perception of information always related to the space and geometry, even when we talk about emotion and personality, we are interested about the pure geometric manifestations (like distance among agents, speeds and densities). Thus, the purpose is to evaluate human perception regarding geometric and cultural features through the following questions: *i)* "Is human perception, regarding geometric features, affected by different camera viewpoints and/or type of avatars?"; and *ii)* "Can people perceive cultural features, in virtual humans, without body and facial expressions?". This content can be seen in Section 4.

[1](Available at: http://rmfavaretto.pro.br/vhlab/)

1.1.3 Perception of *CG* Characters

Although the concept of *Uncanny Valley* in *Computer Animation* is very popular, some questions have been raised: *i)* "Does the exposure to virtual characters, which has been going on for several decades now, reduce the Uncanny effect on people's perceptions?", *ii)* "How the charisma and familiarity with virtual humans correlate to the *Uncanny Valley*?", *iii)* "Does Interactive Environments impact in *Uncanny Valley* effect?", and finally *iv)* "How is our perception impacted if more than one character is presented instead of only one?". These main questions are the motivation for our work which revisits the study of *Uncanny Valley* effects caused by *CG* characters in human perceptions. To try to answer the questions (previously mentioned) related to *Uncanny Valley* theory, we: *1)* recreated Flach et al.[26] research using the same questionnaire containing the same images and videos. With this, we compare the perceptions of the people of seven years ago with the current perceptions regarding the effect of the *Uncanny Valley*. The work of Flach was chosen due to the high number of characters with diverse origins (movies, games, among others). *2)* in the same questionnaire, we include images and videos of more recent characters to evaluate the *Uncanny Valley*. In this case we can observe the effect of the *Uncanny Valley* with these new characters and compare them with the characters from the previous work [26]; *3)* we developed a Virtual Reality (*VR*) application varying the following variables: realism of characters and number of characters. Then, we applied a questionnaire evaluating these characters images and videos, and asked to participants to answer about their feeling while interacting with characters. This content can be seen in Section 5.

1.2 Main Contribution

The performed studies are important in some aspects. First, it is relevant to understand how people's perception works in relation to virtual humans. For instance, which camera point of view is more interesting to better recognize density of pedestrians? While areas as behavioral animation put effort to produce more intelligent and realistic agents, can people have the qualitative assessment of those animations? if we are investing to build a very realistic virtual human, are some aspects that can be considered in order to avoid the Uncanny Valley? This work aims to be a step on the direction of answers to those questions.

1.3 Organization of This Document

This work is organized as follows: Section 2 presents the related works and important researches that inspired in some way the present work, while Section 3 presents the methodology used to simulate, detect and visualize interactions among individuals in video sequences and simulations.

Section 4 presents the proposed techniques to develop a viewer of geometrical and non-geometrical data extracted from groups and crowds; and Section 5 presents details of our research involving perception of CG characters. Finally, Section 6 presents the final considerations of this work.

2. RELATED WORK

This section presents some work related to the concept of perception (Section 2.1), cultural features in crowds (Section 2.2), visualization of features (Section 2.3) and perception of *CG* characters (Section 2.4).

2.1 Perception

According to Schacter et al. [55], perception is important for the representation and understanding of environmental information. Information is captured through the senses and it is identified, organized and interpreted in our brain. One of the ways to study human behavior is through perception. For example, in the case of vision, according to Yantis [60], the selection of visual display information is controlled by at least two distinct ways: *i)* an individual's ability to control which regions or objects in the field of view should be selected for visual post-processing by plotting a set of goals about the current task; and *ii)* stimulus priorities can capture attention regardless of the goals. According to Gregory [30], in addition to receiving sensory stimuli, perception has other processes, such as the involvement of memory. According to Atkinson [4, 5], this perceptive information is passed directly to human memory for later use.

According to Bernstein [6], perception is the process of using information to understand the environment. For example, a person can perceive the geometric shape of a rectangle if he/she already has this information in his/her memory. Also according to Bernstein [6], perception can be divided into two processes: *i)* processing input of sensory information, for example, perceiving the geometric shape and identifying it; and *ii)* processing connected to concepts, which is linked to knowledge, for example, learning a new geometric shape. According to Gregory [30], the brain tries to pre-consciously understand sensory information. However, there are still debates about whether sensory information is sufficient for perception, or whether hypothesis tests are necessary to obtain results from the information captured from the environment.

Regarding the perception of crowds, Lamer et al. [44] conducted a study on the perceptions of interracial crowds. The authors theorized that the rapid visual processing of human crowds provides a means for people to learn distinctions from social categories, being able to differentiate people. In doing so, they tested this theory against racial categorization, with the aim of assessing whether and how the perception of emotionally segregated interracial crowds influences people's racial cognition. In this study, participants were assigned to observe subgroups in interracial crowds in images. The authors separated the subgroups into: control, having people with similar emotions; and emotional segregation, with people with different emotions. Participants exposed to emotionally segregated groups exhibited stronger racial category boundaries. While participants exposed to groups of non-segregated interracial crowds exhibited a stronger racial essentialism, that is, creating stereotypes based on race and not for emotional reasons. Ennis et al. [19] evaluated the effects of camera views and geometric issues (orientation and position) on pedestrian formations. The authors performed a study of how people perceive the characteristics of virtual crowds in static scenes, reconstructed from annotated static images, with changes in the orientations and positions of pedestrians. The authors applied rules based on information from the scene, and from that, they found that the perception of positioned crowds, based on the original pedestrian positions and orientations, was improved compared to crowds with pedestrians having random positions and orientations. In addition, they measured the effect of the camera's point of view on the plausibility of virtual pedestrian scenes, finding that a first-person point of view is less effective than the canonical point of view (which according to the authors is considered a better angle to take a picture) to identify pedestrians with random positions.

Yang et al. [59] conducted a study on analysis of perception to determine the impact of groups at various densities, using two points of view: top and first-person view, both shown in Figure 2.1. In addition to this perception, they analyzed what type of camera position (top view or first-person view) might be best for density perception. The perceptions were obtained through questionnaires answered by users. First, the authors simulated virtual crowds, and then users used the environment to answer the questionnaire with their own perceptions. The authors' results indicated that groups are perceived more in the top view than in the first-person view. Regarding the low densities in the condition of subgroups, the groups were perceived more in the first-person view than in the top view. At medium and high densities, when individuals were shown, groups were perceived more in top view than in first-person view. With regard to density perception, people perceived higher densities in first-person view than in top view. The work of Yang et al. [59] inspired our approach to using different points of view, when visualizing cultural features, in order to obtain people's perception.

Next section discusses work on pedestrians data, other than geometric information such as cultural, emotional and personality aspects.

2.2 Cultural Features in Crowds

This section discusses some work related to pedestrian and crowds behavioral analysis focusing on personality traits, emotion and groups. In particular, we present the cultural features used in our approach, using the *GeoMind Model* [25].

The *OCEAN* [14, 38] is the personality trait model most commonly used for this type of analysis, also referenced as *Big-Five*: Openness to experience ("the active seeking and appreciation of new experiences"); Conscientiousness ("degree of organization, persistence, control and motivation in goal directed behavior"); Extraversion ("quantity and intensity of energy directed outwards in the social world"); Agreeableness ("the kinds of interaction an individual prefers from compassion to tough mindedness"); Neuroticism ("how much prone to psychological distress the individual is") [45]. Durupinar et al. [16] also used *OCEAN* to visually represent personality traits. Visual representation

Figure 2.1: The figure shows the two points of view at a high density. The upper figure shows the top view, and the lower figure the first-person view.

of agents is given in various ways, for example, the animations of the agents are based on two cultural features (*OCEAN* and emotion). If an agent is sad, his/her animation will represent that emotion.

Favaretto et al. [21] proposed a way to detect cultural aspects in crowds based on the *Big-Five* [13] personality model and extract personal behavioral data from video sequences. For this, they took two main steps: video data extraction and cultural analysis. In the first, they obtained individual trajectories of each pedestrian in real videos. In the second step, with data extracted from the trajectories, they did the cultural and personality analyzes. Trajectory data is computed geometrically for each person i in each f frame, thereby obtaining: *2D* positions (in meters), speed (meters / frame), angular variation (degrees). In addition to these data, other data were computed: collectivity, socialization and isolation. Therefore, each pedestrian present in the video had a feature vector with all these mentioned data. Thus, in the second step, such feature vector is mapped to an *OCEAN* dimension vector, with the help of the *NEO PI-R* questionnaire used in the [13] work.

Favaretto et al. [20] also presented a method to detect cultural aspects in groups of individuals, using video sequences. The authors proposed to map some observed features of persons such as speed, distance between them and occupied space, to *Hofstede's Cultural Dimensions* (*HCD*) [34] such as *Power Distance* (*PDI*), *Masculinity/Femininity* (*MAS*) and *Long/Short-term Orientation* (*LTO/STO*). The method is able to identify temporary and permanent group of individuals, the latter been defined if it maintains a group structure for more than 10% of the total frames of the video. The results showed that their defined equations to map cultural aspects seem to be coherent with psychological literature. A similar idea, but using computer simulation and not focused on computer vision, was proposed by Lala et al. [43]. They used *Hofstede's Dimensions* to create a simulated crowd from a cultural perspective. Gorbova and collaborators [28] presented an automatic personality screening system from video presentations in order to decide whether a person should be invited to a job interview based on visual, audio and lexical tips. Dihl et al. [15] generated individual behaviors, also based on *Hofstede's Cultural Dimensions* [34] and applied individual trajectories extracted from video sequences, that is, they did the reverse of the work of Favaretto et al [20]. With this, they generated these behaviors and compared them with the real videos, as shown in Figure 2.2. In this figure it is possible to see in (a) the real image, in (b) the marked pedestrians, and in (c) the generation of the pedestrians.

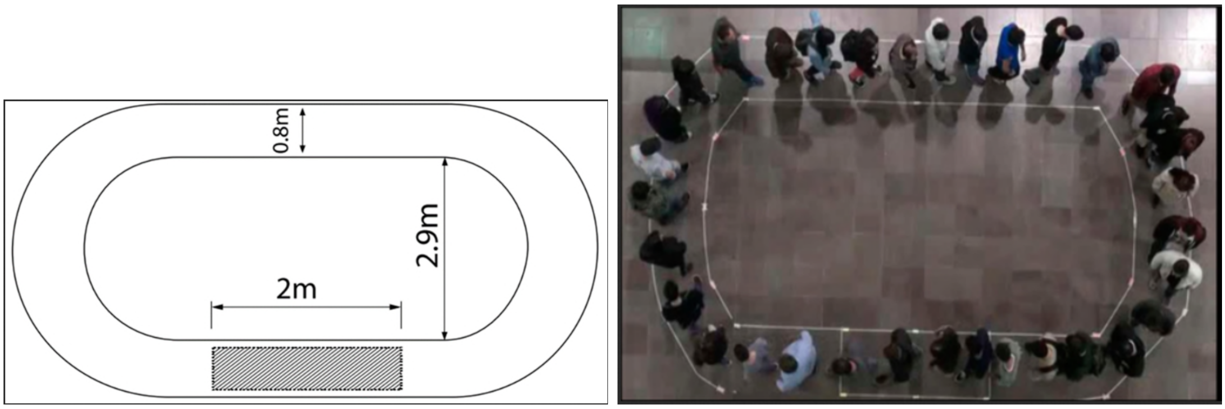
(a)

(b)

(c)

Figure 2.2: In this figure it is possible to see the process of generating behavior from video sequences proposed in [34]. In (a) it is shown the real image, in (b) the marked pedestrians, and in (c) the generation of the pedestrians based on data extracted from video sequences.

In other investigations related to groups, Favaretto et al. [23] investigated cultural aspects in group behaviors (for example, personal space) between Brazil and Germany, using controlled experimental videos, related to the *Fundamental Diagram Experiment* (*FD*) [10] (being experiments between Germany and India). The experiments were carried out with the same number of populations in the *FD*, and can be seen in Figure 2.3.



(a) (b)

Figure 2.3: In this figure it is possible to see the model (a) and the reproduction (b) of the *FD* [10] experiment made in the work of Favaretto et al. [23].

Several models have been developed to explain and quantify basic emotions in humans. One of the most cited is proposed by Paul Ekman [17], who considers the existence of 6 universal emotions based on cross-cultural facial expressions (anger, disgust, fear, happiness, sadness and surprise). In another work by Favaretto et al. [24], the authors proposed a way to detect pedestrian emotions in videos, based on *OCC Emotion Model*. The *OCC* (Ortony, Clore, and Collins) emotional model indicate the susceptibility of each of the five personality factors to feeling every emotion. To detect the emotions of each pedestrian, the authors used *OCEAN* as inputs, as proposed by Saifi [54], and they proposed an emotional mapping for each of them, as shown in Table 2.1. In this table you can see that there are four emotions (fear, happiness, sadness and anger) being mapped by the *OCEAN* factors. In our approach, we proceed with an analysis to assess if participants can perceive geometric features, as well as emotions and personalities in scenes when pedestrians are represented by avatars.

Table 2.1: *OCEAN* mapping for the four emotions: fear, happiness, sadness and anger.

OCEAN Inputs	**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

Next section presents some discussion regarding visualization of features.

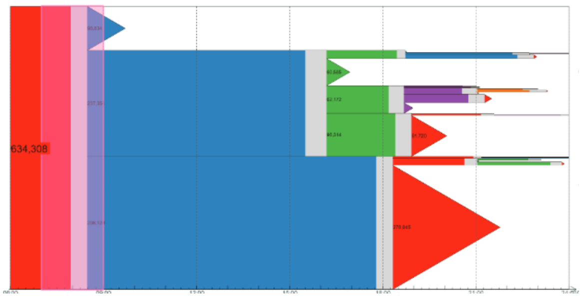
2.3 Visualization of Features

This section presents work related to the visualization of cultural features in crowds. The methods presented in this section were important for inspiration on the visualization methods developed in this research.

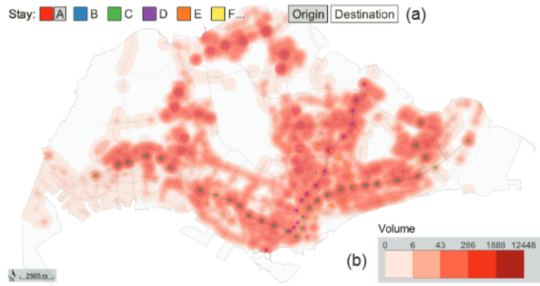
The work of Zeng et al. [62] deals with the rhythms of the daily movement of large crowds of humans. They define rhythms as the trajectories that each crowd traverses along the day (for example, home-school-home). One of the applications of their research is to help public transport companies to understand which trajectory is more crowded and what can be done to improve the crowd flow. This method proposes to analyze such data through interactive visualizations and to allow the users to explore them. The used data are from urban public transport in Singapore. Users can manipulate data and views, which are divided into three types: *i)* Sequential visualization of rhythms; *ii)* Visualization of the density of the rhythms; and finally *iii)* Statistical view of rhythms. The first visualization option, which can be seen in Figure 2.4(a), presents the rhythms in a tree structure, where each color represents a location, for example, the red-blue-red color sequence means that the crowd left a place represented by the color red, passed a blue place, and returned to red place, thereby representing a rhythm. This view also features a time controller to view crowd time in each color, and the user can manipulate this controller and set this time. In Figure 2.4(b), the second view is shown, which shows the density of each location through a heat map and it is possible to set the pace and density of this rhythm. The third view is shown in Figure 2.4(c), where the percentage of each rhythm is illustrated through a bar chart, i.e., showing which rhythm occurred most frequently.

The work of Ardeshir and Borji [3] shows experiments and graphs made between two points of view (first-person and top cam view), thus helping in the integration and use of the types of cameras used in the present work. The graphs can be seen in Figure 2.5(a), where the left graph vertices are given by the view of each pedestrian present (first-person view), while the right graphs are given by the top view, where each image has the number of pedestrians present in the video, and each vertex is the representation of each pedestrian. Still in this figure, the graphs are also interconnected, each vertex on the left is represented by a graph on the right, and this is done to be able to relate the two types of cameras present. Figure 2.5(b) shows that in a top view you can see the first-person views present in the current frame.

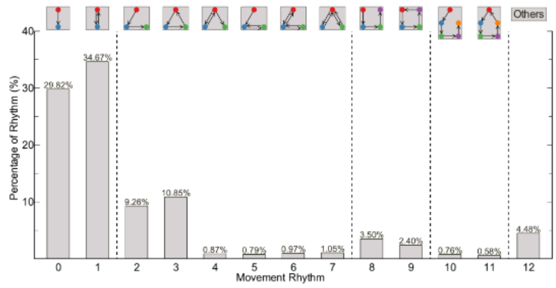
Although far from the goal of this work, Zhao et al. [63] introduced *PEARL*, a visualization tool for analyzing personal emotion styles on social media. Important to mention that this kind of application has been very explored in literature, nowadays. We chosen to mention only one in order to show another aspect of visualization of emotions. In this case, the tool allows the extraction of textual data from people on social media to interactively analyze their emotions over time and get their emotional styles. These styles are defined through tweets published by people, and the more publications the more the flow in the view increases, as shown at the bottom of Figure 2.6, where it has the lighter blue. In this Figure 2.6, (a) shows the distribution of emotions over time



(a) Sequential visualization of rhythms

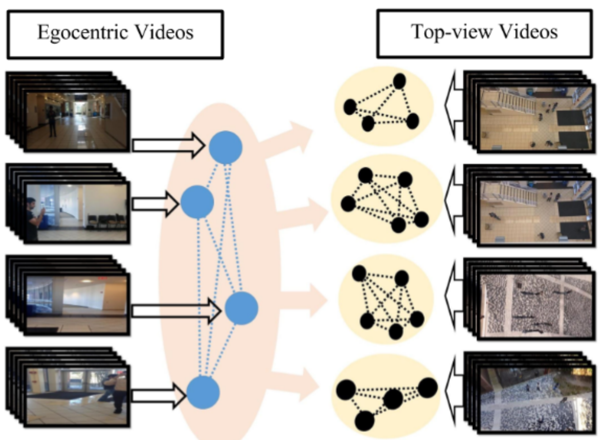


(b) Visualization of the density of the rhythms

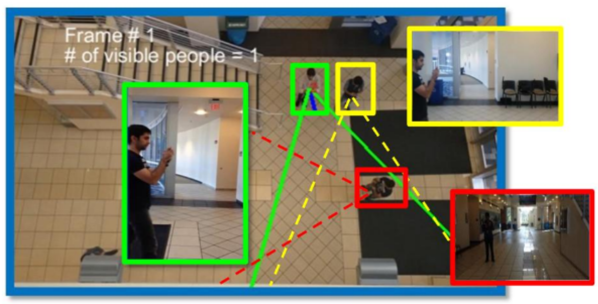


(c) Statistical view of rhythms

Figure 2.4: The figures present rhythms, or trajectories, of crowds. Figure (a) presents the rhythms in a tree structure, where each color represents a location, for example, the red-blue-red color sequence means that the crowd left a place represented by the color red, passed a blue place, and returned to red place, thereby representing a rhythm. In (b), the density of each location is shown through a heat map and it is possible to set the pace and density of this rhythm. In (c), percentage of each rhythm is illustrated through a bar chart, i.e., showing which rhythm occurred most frequently.



(a)



(b)

Figure 2.5: In this figure it is possible to see the graphs representing the views of the types of cameras present in the work of Ardeshir et al. [3]. In (a) first-person graphs and top graphs. In (b) it is shown that in a top view, it is possible to see the first-person views present in the current frame.

and their coding, i.e., how much emotion appears over published tweets. In (b) it is possible to see the degree of positivity of each emotion in the tweet through a set of arrows, which are activated when the mouse cursor goes over the figure indicated in (c). In (d) it is possible to see all emotions and emotional states through captions, and being represented in (a) each by a different color. In addition, this viewer still allows the viewing of published words and tweets, this can be done with the mouse click on the part where has the lightest blue. With this, in this Figure 2.6 it is possible to notice that this analyzed person is emotionally resilient since he was able to recover quickly from (and) his negative emotional states.

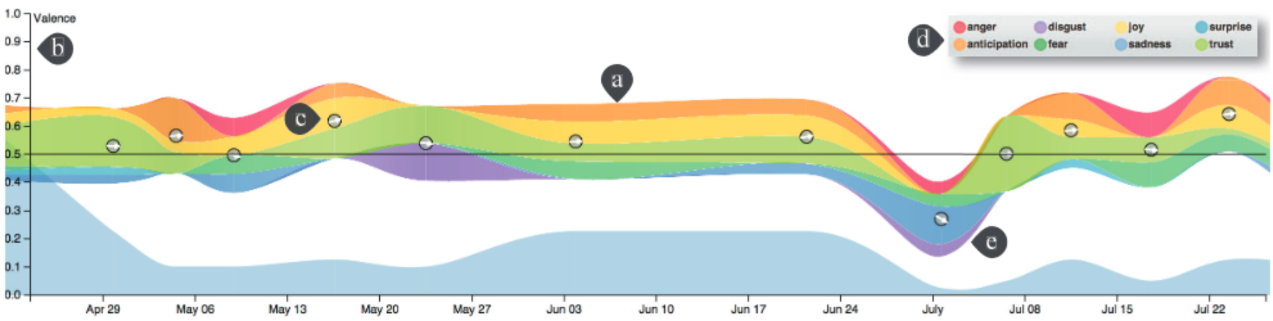


Figure 2.6: PEARL, a visualization tool for analyzing personal emotion styles on social media, created by Zhao et al. [63]. In (a) it is shown the distribution of emotions over time and their coding. In (b and c) it is shown the degree of positivity of each emotion in the tweet. In (d) it is possible to see all emotions and emotional states through captions, and being represented in (a) each by a different color.

Until here we are discussing papers where pedestrians positions, other geometrical information, extracted cultural aspects or still text, as in the last example, are the focus of the studies, but without considering faces animation. Next section aims to discuss some research on perception performed in the context of CG, considering animated faces.

2.4 *CG* Characters and Uncanny Valley

This section discusses some work related to the analysis of *Uncanny Valley* effects caused by *CG* characters. The *Uncanny Valley* is a theory created by roboticist Masahiro Mori [49] who analyzes the emotional reaction of humans to artificial beings. According to this theory, if robots have a high degree of realism close to humans, they may fall into the "*Uncanny Valley*", which can cause an eerie impression on the viewer. Thus, when robots show signs of life, such as movement or conversation, it tends to generate high peaks and valleys, changing the shape of the *Uncanny Valley*. Figure 2.7 shows the *Uncanny Valley* chart created by Masahiro Mori [49]. Please, see [49] for further details about the research.

Inspired in Mori's work, several other researchers have used/study *Uncanny Valley* effect to measure the discomfort of artificial characters (robots, characters created with *Computer Graphics*, among others). For example, in the work of Katsyri et al. [39], the authors reinterpreted the original *Uncanny Valley* hypothesis and revisited empirical evidence for theoretically motivated *Uncanny Valley* hypotheses. Thus, that work helped to understand and form the axes of the *Uncanny Valley* charts (X-axis being the human likeness, and the Y-axis being the comfort) in this present work.

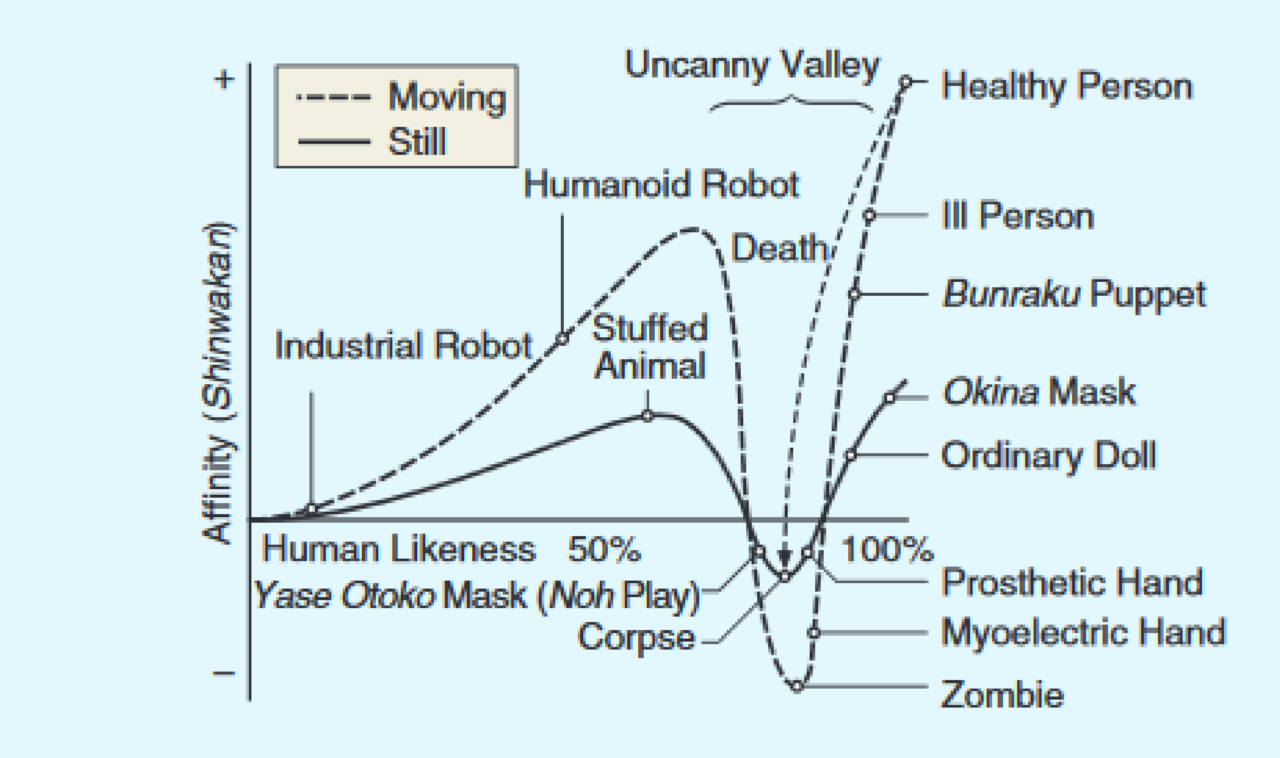


Figure 2.7: *Uncanny Valley* chart created by Masahiro Mori [49].

In the work of Zell et al. [61], the authors analyzed two factors that define how a character looks like: shape and material. With the help of artists, they designed a set of elaborate stimuli, consisting of different levels of stylization for both parameters, and analyzed how different combinations affect the perceived realism: appeal, strangeness, and familiarity of the characters using *Uncanny Valley* effects. In addition, authors investigated how such combinations affects the perceived intensity of different facial expressions.

The effect of the *Uncanny Valley* hypothesis on human perceptions of *3D* models has been investigated in *Computer Graphics*. One of these studies is the work of MacDorman and Chattopadhyay [47], where the main objective was to determine whether reducing the consistency of realism in visual characteristics increases the effect of the *Uncanny Valley*. The hypotheses of the authors are based on the theory of the inconsistency of realism, which predicts that the effect of the *Uncanny Valley* is caused by an entity that has characteristics, and not all are perceived as belonging to a real living anthropomorphic being. With this hypothesis, they investigated the animacy for humans and non-human animals, and realism for humans, non-human animals and non-human objects.

Based on facial capture techniques and technologies, Seymour et al. [57] studied the interactive avatars effects using photo-realistic human faces. Based on recent advances in real-time rendering technology, the authors observed the effect of the *Uncanny Valley* theory on user interaction with the photo-realistic human avatar. Following the theory proposed by Mori [49], that the movement amplifies the "strange" effect, the hypothesis of the authors is that the interactivity increases even more this strangeness.

The work of Hodgins et al. [33] and Hyde et al. [35] reported the importance of the realism of characters created with *CG*. In the first, the authors performed perceptual experiments, exploring the relative importance of different anomalies using two methods: a questionnaire to determine the emotional response to complete vignettes, with and without facial and audio movement; and a task to compare the performance of a virtual "actor" on short clips (extracted from the vignettes) representing a series of different facial and body anomalies. In the second, the authors conducted two experiments showing how exaggerated facial movement influences the impressions of cartoons and more realistic animated characters.

Flach et al. [26] investigated the *Uncanny Valley* theory to evaluate their effects on the perception of *CG* characters used in movies, animations and computational simulations. The authors evaluated the human perceptions about these characters through a questionnaire containing images and videos of these characters, to obtain answers to the following research questions: "Does the *Uncanny Valley* exist in *CG* characters?" and "Does adding movement to these characters change the shape of the *Uncanny Valley* curve, like Mori suggested?". In present work, we recreated the Flach et al. [26] experiment, with the same questionnaire and the same images and videos of the *CG* characters, which is shown in Figure 2.8.

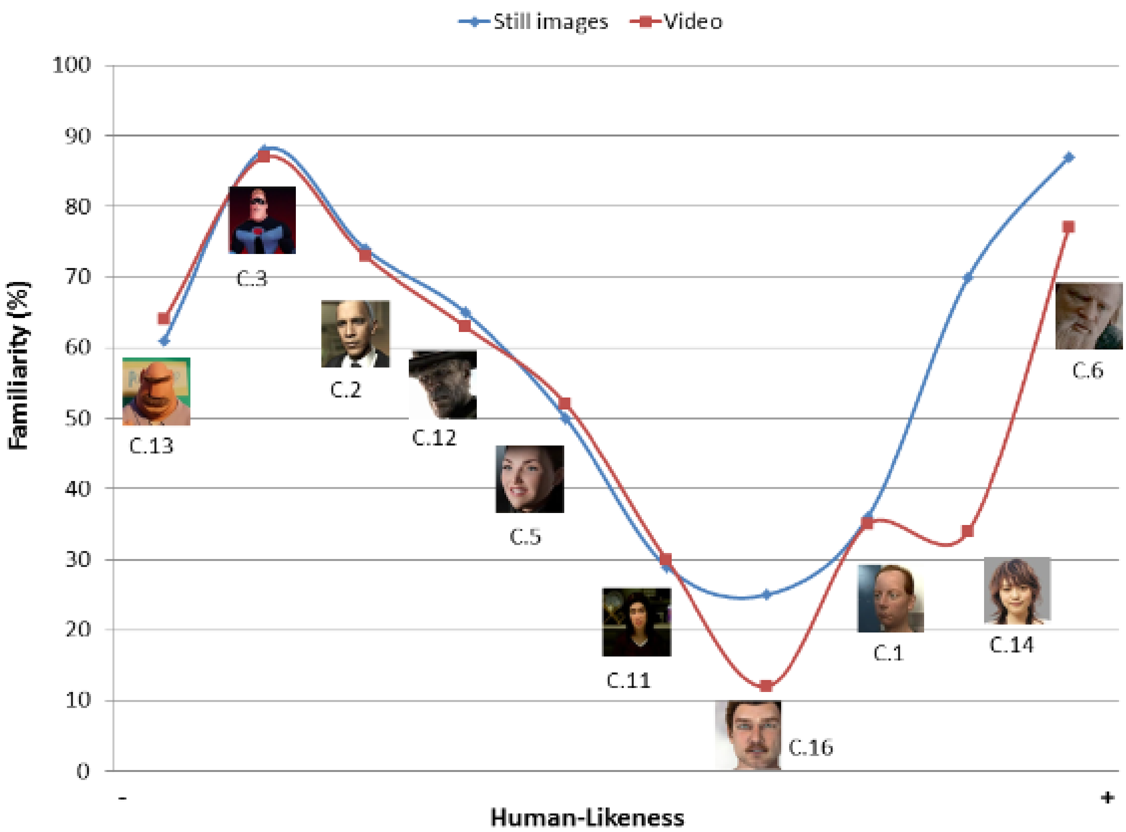


Figure 2.8: *Uncanny Valley* chart from the work of Flach et al. [26]. The Y-axis represents *Familiarity* (in our approach we call *Comfort*, which is explained in Section 5) with the character. The x-axis, on the other hand, represents *Human-Likeness*, indicating from left to right the most realistic characters, that is, those who look more like human beings.

Next sections describe the three investigations performed in this work. Firstly, Section 3) is concerned with the visualization of interaction in crowds as a way to have a qualitative assessment of those events. Then, we present the developed viewer to provide visualization of geometric and non-geometrical aspects of crowds, as emotional status (in Section 4). Finally, Section 5 presents our investigation regarding animated faces.

3. PERCEPTION OF PEOPLE INTERACTIONS IN CROWDS

This section aims to present the work done on interactions between virtual agents in crowd simulations and between individuals in real crowd videos. This section is divided into two sections: Section 3.1, where our methodology to detect and model interactions is explained, and Section 3.2, where the results on this content are presented.

3.1 Methodology of Perception of People Interactions in Crowds

As already mentioned, a common phenomenon between two or more individuals is the interaction (for example, communication between them). Only a few of the existing simulation methods take this phenomenon [42, 11] into account to generate simulated scenarios. Interaction among agents is an important feature in order to provide realism in games and movies, once real people interact in real life. In addition, it can be used to find out patterns of behaviors or events in the simulations. In this case, we considered that the possibility to visualize data, generated by simulations with interactive agents, can be useful to understand the performed simulation and behaviors behind the game. In this context, individuals can interact multiple times, anytime and anywhere. To have physical interaction among people, they need to be aware of their environment (for example, through vision), and must want to interact. Therefore, in order to have data set containing interactions among individuals with ground truth, we developed a simple model to simulate interactions in groups simulation. So, this section is divided into four parts: Section 3.1.1 defines our interaction model, Section 3.1.2 describes how we define personality traits for agents using the *OCEAN Model* [27], Section 3.1.3 shows how the mode of interaction between people in real videos was defined and finally Section 3.1.4 presents the methods used to visualize the interaction data.

3.1.1 Simulating Interactions using *BioCrowds*

In order to simulate crowd of agents [1], we chose to work with the simulation method BioCrowds [7], since it is a state-of-the-art simulation technique which guarantees a free-collision movement for agents. It is based on Runions [53] spatial colonization algorithm adapted to crowds. To perform this adaptation, some changes are proposed by Bicho[7]:

- **Restricting auxin space**: only auxin contained in the agent's personal space can influence its movement;
- **Auxins Persistence**: auxins are kept in the virtual environment during the simulation, but are available only to the nearest agent. This distance calculation is updated every iteration;

[1] Work developed in collaboration with Paulo Knob, PhD student and colleague at VHLAB.

- **Goal Seeking**: Besides being influenced by auxins, the movement of individuals is also influenced by the willingness of each individual to reach a particular destination; and

- **Speed Adaptation**: agents vary their speed according to space availability.

In our proposed model, we add a generic way to provide interaction among agents. To do so, we create an Interaction Factor γ for each agent, where $\gamma = [0, 1]$. This value represents the willingness of an agent to interact with other agents, where a high value (i.e., 1) represents a great will to interact and a low value (i.e., 0) represents a small will to interact. In our method, an interaction between two agents starts when they are close enough, simulating a spatial perception. However, before the interaction occurs, we make it possible for agents to call attention of each other, so they can start to approach instead keep going to their respective goals. In this matter, we use the concept of Personal Space defined by Hall [32] which represent the relationship among individuals, where intimate space is characterized by maximum distance of $0.45m$, personal is $1.2m$, social is $3.6m$ and public is $7.6m$. Following this concept, we define a threshold $\zeta = 7.6$, which is used to define the distance where agents can call attention of each other and it is defined as the Public Space from Hall (i.e., $7.6m$). Therefore, agents can call each other attention, if:

- The distance between the two agents is smaller than $7.6m$ at frame f: $(Dist(\vec{X_a^f}, \vec{X_b^f}) < \zeta)$, where $Dist$ states for the Euclidean distance between two vectors and a and b are agents;

- The interaction factors γ_a and γ_b of both agents are higher than a random value Rv^f.

Such random value Rv^f is generated, at frame f, when $\delta(\vec{X_a^f}, \vec{X_b^f}) < \zeta$, for each agent involved and tested against their respective interaction factors. So, for example, if $Rv_a^f \leq \gamma_a$ and $Rv_b^f \leq \gamma_b$, they call each other attention and start to move towards each other (and then Rv^f keeps fixed). If an agent fails this test, they do not call attention of each other, consequently they do not approach to each other, so maybe the condition regarding the distance is going to fail and this pair of agents are not going to interact. Concerning the range values for Rv^f, we have made it between 0.1 and 0.5 to guarantee that high values of γ generate interactions.

Once two agents called each other attention, they start a new interaction group and begin to move towards their group center position $(\vec{X_g^f}))$, i.e., the center position between the pair of agents at frame f. So, agents which do not call attention to the other can not interact. While they are approaching the group center position, their speeds are reduced according to their distance to such goal, so the closer an agent is of the center of its interaction group, the slower it walks. We do so to avoid agents walking at high speeds, while approaching each other, and stopping suddenly, which would not be natural. To do so, we use the formulation presented in Equation 3.1 to provide the speed reduction in agent a:

$$\beta_a^f = \sqrt{(Dist(\vec{X_a^f}, \vec{X_g^f}) - \omega)/(\zeta - \omega)}, \quad (3.1)$$

where $Dist(\vec{X_a^f}, \vec{X_g^f})$ is the distance between the agent a and the center of its respective interaction group g, in a given frame f. The threshold ω is used to define the distance where agents can interact and it is defined as the Personal Space according to Hall (i.e., $1.2m$). The speed reduction β_a^f assumes a value between 0.3 and 1 and represents a percentage of the desired speed an agent will assume at a given frame f, so $\beta_a^f = [0.3, 1]$. We chose such interval of values to avoid agents walking too slow, so we clamp the reduction in 30% of the desired speed. In a similar way, β is computed for all agents in interaction situation.

When two agents (a and b), that were approaching to each other, reach the threshold ω, they may stop moving and start to interact with each other, if:

- The distance between these two agents is lower than $1.2m$ $(Dist(\vec{X_a^f}, \vec{X_b^f}) < \omega)$;

While two agents interact, they keep together in a certain distance and certainly their γ values are greater than Rv^f (as showed in last section as a condition to interaction happens). So, if nothing changes to update such variables, they could interact forever. To solve this, we propose a decay function of γ value as time passes, if the agent is involved in some interaction. In other words, if two agents are interacting, their respective interaction factors, γ_a and γ_b, start to reduce at each frame by a constant Ω, as follows:

$$\gamma_a = \gamma_a - \Omega, \tag{3.2}$$

where $\Omega = 0.05$ (empirically defined). Therefore, as time passes by, agents lose interest to interact and, eventually, follow their respective goals.

It is interesting to notice that, although both process (i.e., call attention and interaction) occur between two agents, a larger group of agents can be calling attention or interacting at the same time, pair by pair. For example, an interaction group can have three agents, where agent a interacts with agent b, agent b interacts with agent c and agent c interacts with agent a. Also, the interactions occur in both ways, so if agent a is interacting with agent b, agent b is also interacting with agent a.

Another important issue is to allow agents that already interacted, to interact again. As explained previously, agents decrease their interaction factors as they interact with each other. So, it is possible that an agent starts the simulation with a high interaction factor, interacts with some other agent and leaves this interaction with a low γ value, making it not able (or at least, most unlikely) to interact again. To solve this, we define a constant $\beta = 150$, which represents the time (in frames) an agent takes to recover its original interaction factor value. So, after β frames that an agent stops to interact, its interaction factor γ is reseted to its original value, allowing it to interact again.

3.1.2 Defining Agents Personalities

The interaction factor γ can be statically defined by the user or even randomly generated for each agent. Although, this analysis also proposes to define this factor as a function of a personality input. To do so, we chose to work with the OCEAN (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism) psychological traits model, proposed by Goldberg [27], since it is the most accepted model to define an individual's personality. Therefore, each agent in our simulation have defined values for OCEAN traits. Following such psychological method, we should define how each OCEAN factor would affect an individual's will to interact. To do so, we take into account the definition of each factor, in short:

- Openness (O): reflects the degree of curiosity, creativity and a preference for novelty and variety;
- Conscientiousness (C): reflects the tendency to be organized and dependable, preferring planned action than spontaneous behavior;
- Extraversion (E): reflects the sociability and talkativeness;
- Agreeableness (A): reflects the tendency to be cooperative and compassionate with others; and
- Neuroticism (N): reflects the tendency to experience unpleasant emotions easily and the degree of emotional stability.

Following the previous definition, we determine the relationship between each OCEAN factor with the willingness of the agent to interact, as well the impact of each of those factors on it. Table 3.1 shows how OCEAN relates with the interaction factor. A positive relationship means that the higher the factor, the higher the interaction factor γ is too (and vice-versa). A high impact means the factor is very important to determine the interaction factor γ (and vice-versa).

Table 3.1: Relationship between each OCEAN factor with the willingness of the agent to interact. A positive relationship means that the higher the factor, the higher the interaction is too (and vice-versa). A high impact means the factor is very important to determine the interaction level (and vice-versa).

Factor	Will to Interact	Impact
O	Positive	High
C	Negative	Low
E	Positive	High
A	Positive	Low
N	Negative	Low

Therefore, the interaction factor γ_a for each agent a is defined as follows:

$$\gamma_a = (W_h O_a) + (W_l(1 - C_a)) + (W_h E_a) + (W_l A_a) + (W_l(1 - N_a)), \tag{3.3}$$

where each OCEAN factor value lies between [0,1] and W_h, W_l stands for the High and Low impacts, respectively. These values were empirically defined in this analysis as $W_h = 0.45$ and $W_l = 0.05$.

3.1.3 Interactions in Video Sequences

While in last section we simulate interaction behaviors among agents in a crowd simulator, in this section we are interested in the detection of interactions among pedestrians in real video sequences and consequently their visualization. We use the Cultural Crowd data set proposed by Favaretto et al. [21] that contains videos from various countries, and we use *GeoMind Software* [25][2] to generate files with data extracted from each person present in the videos. The data generated for each person are: position (X_i, Y_i) of each person i in a given video (already in the world coordinates) at each f frame; distances between i and all other individuals, which will be used to determine if they are interacting or not. In addition, *GeoMind* [3] extracts OCEAN personality traits of each individual, which are also used in this analysis.

In the work proposed by Favaretto et al. [21], the authors present a methodology to detect the OCEAN personality traits. Based on filmed sequences, pedestrians are detected, tracked and characterized. Such information is then used to find out cultural differences in those videos, based on the Big-five personality model. For this, they used the NEO PI-R [13] that is the standard questionnaire measure of OCEAN Model. Firstly they selected NEO PI-R items related to individual-level crowd characteristics and the corresponding factor (for example: "Like being part of crowd at sporting events" corresponding to the factor “Extraversion”) and then propose a way to map this data extracted from video sequences to OCEAN parameters, generating values for O, C, E, A and N of all individuals. Then, we use the same formulation defined in Equation 3.3 to calculate the interaction factor γ for each person detected in the video. Once we have the distance between each person, for each frame, and the interaction factor γ for each person in the video sequence, we can define the interaction between person i and person j following the same two conditions, as described in previous sections:

- The distance between person i and person j must be less than 1.2m ($Dist(\vec{X_i^f}, \vec{X_j^f}) < \omega$);
- The interaction factor γ of both individuals is higher than a random value (explained in previous sections).

[2] Developed by Dr. Rodolfo Migon Favaretto.

[3] *GeoMind* generates other parameters that are not used in this analysis. For further details, please see [25]

3.1.4 Interactive Visualizations Methods

The visualization of large amounts of data is essential to data understanding. Not choosing a suitable technique may generate confusion or misunderstanding. The data set generated by the interactions simulations, as described in lasts sections, has a large amount of data that needed to be well defined and treated. Such data could be displayed using several visualization techniques, therefore, four methods were chosen: bar chart, network graphs, scatter plot and time-line chart with slider.

The bar chart was chosen due to the need for quantitative analysis of the interactions by simulation and video. In this visualization, the X-axis represents each simulation or video analyzed, while the Y-axis presents the amount of interactions for each of them. The size of the bar is given by the amount of interactions that each simulation or video had. Indeed, the amount of interactions was divided by two, due to the fact that an interaction between two agents/individuals is bidirectional, it means, if agent/person a interacts with agent/person b, agent/person b also interacts with agent/person a. Figure 3.1 presents such visualization.

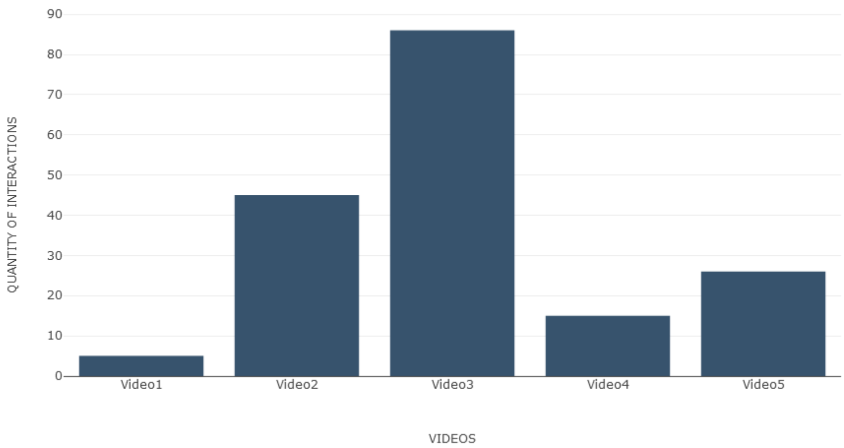


Figure 3.1: Bar Chart Example. The X-axis shows examples of videos, while the Y-axis shows the number of interactions for each video.

The scatter plot visualization method was chosen to show the number of interactions in the physical space. In this visualization X-axis and Y-axis represent the exact position of an agent/person, in the environment, that was interacting in a given simulation or video sequence. For example, if an agent/person was interacting at the position (5,10), such interaction is shown in the visualization at $X = 5$ and $Y = 10$. The number of interactions happened in a certain position in the space is represented by the circle size. Also, it is possible to change the simulation or video visualized at the moment. Figure 3.2 presents such visualization.

The network graph was chosen to provide a visualization method that aims to demonstrate the relationship between the interacting agents/individuals at each simulation or video sequence. Here, each agent/person is represented as a node. Each node can be connected by edges with other nodes, where each edge represents a relationship between these two nodes (i.e., two

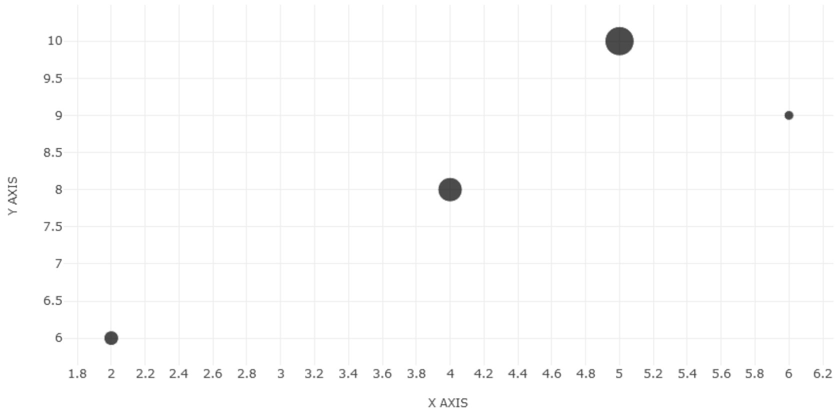


Figure 3.2: Scatter Plot Example. This chart shows the number of interactions in each 2D position. The size of the circle represents the number of interactions, that is, the more interactions the larger the circle. For example, the position circle (5, 10) is larger than the others, as it was the position that had the most interactions.

agents/individuals interacted). The more relationship a node has, the closer to the center it will be in the visualization. The size of each node is given by the amount of interactions for that agent/person. In the case of the video sequences, since there are no fixed OCEAN values, no color is assigned. As it was already done in the previous visualization, it is possible to change the simulation or video visualized at the moment. Figure 3.3 presents such visualization.

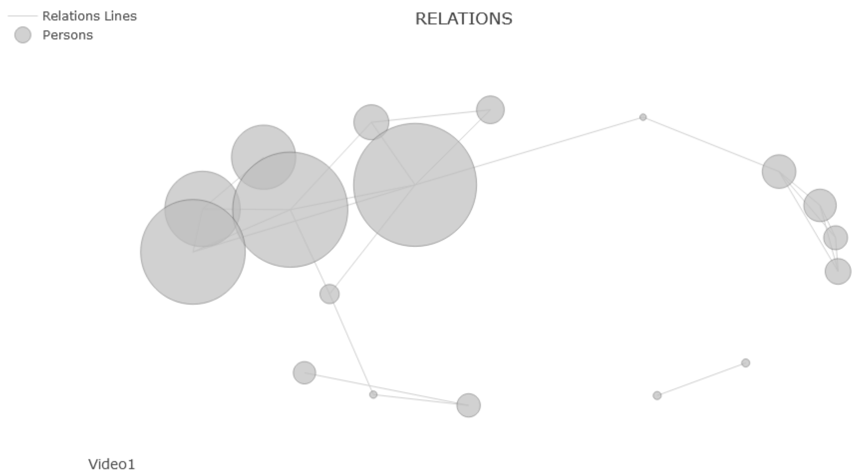


Figure 3.3: Network Graph Example. This figure presents a graph of relations, in which case each node represents a person, the size of the node is represented by the amount of interactions that person has had, and each edge represents a relationship between two individuals.

The time-line chart was chosen due to the need to represent the number of interactions by the frames of each simulation or video sequence. This method is essential to visualize temporal data. Each frame has a number of interactions. The X axis shows the frames of the set, while the Y axis presents the number of interactions. In addition, this method uses a slider to facilitate filtering between frames. With the slider, it is possible to select a range from an initial frame to a final frame. The number of interactions per frame is shown in the chosen range. As it was already done in the previous visualization, it is possible to change the simulation or video visualized at the moment. Figure 3.4 presents such visualization.

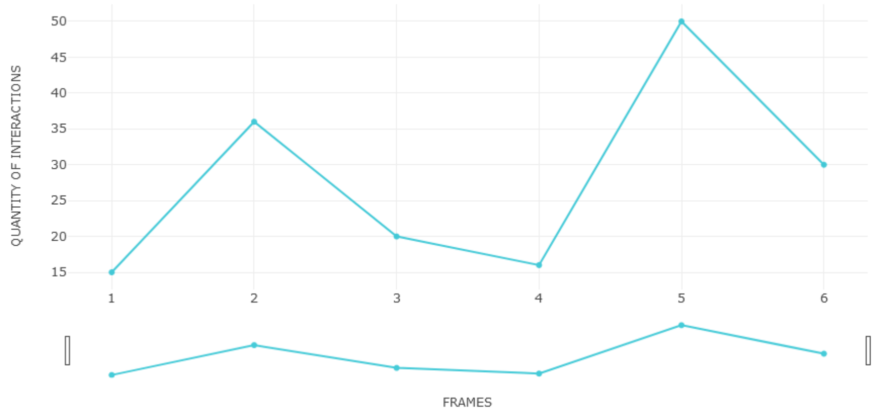


Figure 3.4: Time-line Chart Example. This chart shows the number of interactions across the frames. In addition, it has a time controller to choose to view in a certain period of time.

To build our visualizations, we chose to work with *Plotly*[4]. *Plotly* is a library for *Python* and other languages (*JavaScript*, *R* and etc.) that provides visualization tools. *Dash*[5] is a framework that helps in web development of visualization applications. The development is all done within a Dash application, where you can create HTML "Divs" that help you to visualize the *Plotly* charts, buttons, tabs, captions, among other options. In addition, Dash allows functions for changing the views data and the options cited.

Further results regarding visualization of interactions are described in Section 3.2.

3.2 Results of Perception of People Interactions in Crowds

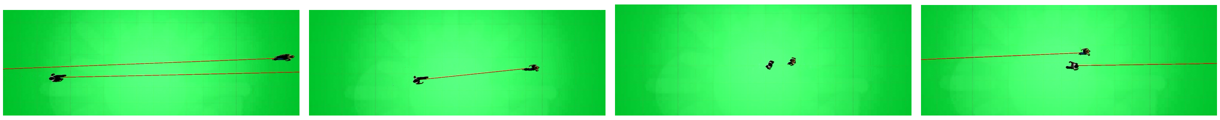
This section presents some results achieved by the performed analysis on perception of interactions in crowds. Section 3.2.1 shows a briefly evaluation of the method developed to simulate interactions among agents. Section 3.2.2 shows how we generated all simulation data for the visualizations, while Section 3.2.3 shows how we generated the interaction data from video sequences. Section 3.2.4 shows the visualizations we built.

3.2.1 Interactions Simulation

To assess the proposed model, it was first important to verify that the model was performing the expected crowd behavior. To do this, we used the method explained in Section 3.1.1 to simulate interactions between agents in a simple environment. We modeled a 30x10 scenario with

[4]Plotly is available at https://plot.ly/
[5]Dash is available at https://plot.ly/dash/

two goals and two agents. Each goal is placed on one side of the environment (that is, one on the left side and the other on the right side), and each agent spawns at one goal position and wants to reach the opposite goal (that is, an agent starts at the first goal position, on the left, and wants to reach the second goal, on the right). The other agent performs the opposite.



(a) Walking (b) Calling Attention (c) Interacting (d) Leaving

Figure 3.5: Simple scenario to test our interaction method. Agents generally trying to achieve their respective goals (a). In due course, agents will "call" each other's attention and begin to approach (b). When close enough, they begin to interact (c) and remain as long as the γ interaction factor is high enough. When γ is too low, they stop to interact and follow their respective paths (d).

As shown in Figure 3.5, agents generally walk around, trying to achieve their respective goals (Figure 3.5a). In due course, agents begin to "call" attention to each other and then begin to approach (Figure 3.5b). When they are close enough, they begin to interact (Figure 3.5c) and remain as long as the γ interaction factor is high enough, following the model presented in Section 3.1.1. When γ is too low, they stop to interact and follow their respective paths towards their goals (Figure 3.5d). Therefore, we considered that our simulation method, to provide interactions between agents, seems to work as intended.

3.2.2 Data Generation from Simulations

To generate different simulations in which agents can interact, we model a 30x30 scenario. As the goal-seeking behavior is irrelevant to this analysis, we just start agents trying to reach a random position in the environment. When they reach, a new random position is generated, an so on. As explained in Section 3.1.1, agents can interact with each other, depending on their parameters. We generated 18 simulations, varying the number of agents and the OCEAN input, as shown in Table 3.2. The idea is to verify how the agents would behave with three different OCEAN inputs: a Neutral Personality, a Blue Personality (for example, a pessimist/negative individual) and a Pink personality (for example, an optimistic/positive individual). In addition, we also evaluate the impact of the number of agents on results. These personalities were chosen following the concept of emotion discussion in the personalities, as observed in literature [13]:

- 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-: person known for emotional stability;
- N+ : person feels negative emotions;

Pink agents are expected to be more spontaneous and try to interact more with other agents. On the other hand, blue agents are expected to be more introverted and try to avoid interaction with other agents and just follow their respective paths. The OCEAN values used as input for each personality are defined as follows:

- Neutral personality: O $= 0.5$, C $= 0.5$, E $= 0.5$, A $= 0.5$, N $= 0.5$
- Blue personality: O $= 0.2$, C $= 0.2$, E $= 0.2$, A $= 0.2$, N $= 0.8$
- Pink personality: O $= 0.8$, C $= 0.8$, E $= 0.8$, A $= 0.8$, N $= 0.2$

Table 3.2: Data Generation Simulations.

Sim. number	Qnt Agents	Blue Person	Pink Person	Neutral Person
1	10	100%	0%	0%
2	10	0%	100%	0%
3	10	0%	0%	100%
4	10	50%	50%	0%
5	10	75%	25%	0%
6	10	25%	75%	0%
7	50	100%	0%	0%
8	50	0%	100%	0%
9	50	0%	0%	100%
10	50	50%	50%	0%
11	50	75%	25%	0%
12	50	25%	75%	0%
13	100	100%	0%	0%
14	100	0%	100%	0%
15	100	0%	0%	100%
16	100	50%	50%	0%
17	100	75%	25%	0%
18	100	25%	75%	0%

As can be seen in Table 3.2, three different values were used for the number of agents: 10, 50 and 100 and three different personalities: pink, blue and neutral. We also combine these personalities to analyse the effects. We expect that simulations with pink agents will generate more interactions than simulations with neutral ones, which should generate more interactions than simulations with blue personality, assuming the same number of agents. Furthermore, we expect that simulations with more pink agents will generate more interactions than simulations with more blue, assuming the same quantity of agents. Finally, we expect that more agents in the simulation will generate more interactions. Each simulation is run for 6000 frames. When each simulation ends, a file is generated with information about the position of the agents in each frame and their respective interactions, also in each frame. These files are used in Section 3.2.4 to generate our set of visualizations.

3.2.3 Data Generation from Video Sequences

To generate the results of the interactions in the video sequences, we followed the method explained in Section 3.1.3. Table 3.3 shows some of the videos from the original data set (*Cultural Crowds*), with information about quantity of individuals and amount of frames. [6] All data found for these videos are added to a new data set file, which follows the same structure as the data set generated by the simulations. Therefore, we imported such data set into our visualizations.

As can be seen in Table 3.3, the video sequences have a varied amount of individuals and duration time. The value of the OCEAN feature is calculated for each person for each video, as previously explained. As expected in the simulations (Section 3.2.2), we expect that individuals with OCEAN values more similar to those defined as Pink personality are those who interact most, while individuals with OCEAN values similar to those defined as Blue personality are those who interact least with each other. In addition, we expect that the higher the amount of individuals in the video, the more they are likely to interact. These files are used in Section 3.2.4 to generate our set of visualizations.

3.2.4 Visualizations of Interactions in Crowds

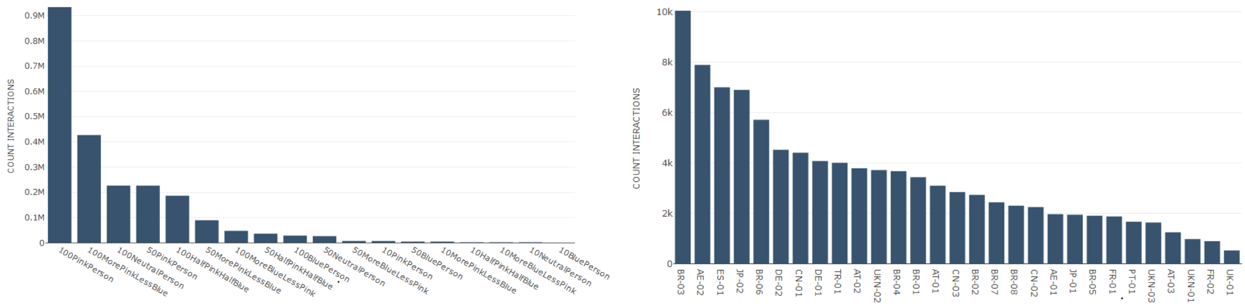
The first visualization option was the bar chart, where we show the number of interactions for all simulations and video sequences. Both can be seen in Figure 3.6. The X-axis is given by a list of simulations/videos, and the Y-axis is given by the amount of interactions. In Figure 3.6(a), we have Simulations X Interactions. In this visualization, as we expected, we can see that the simulation with more agents with Pink personality (100PinkPerson) was the one that generated more interactions, while the simulation that generated less interactions was the only one with Blue personality agents (10BluePerson). In addition, this visualization made it possible to perceive an interesting result. The 100BluePerson simulation, which contains only agents with Blue personality, had more interactions than the simulation 10PinkPerson, which contains only agents with Pink personality. This suggests that, in fact, the number of agents affects the agents' interaction behavior.

[6] We did not present individual OCEAN values due to the amount of information

Table 3.3: Video sequences data from *Cultural Crowds*. [21] data set.

Video ID	Qnt People	Qnt Frames	Time (seconds)	Average γ
AE-01	12	119	5.17	0.53
AE-02	23	229	9.95	0.36
AT-01	12	338	14.69	0.67
AT-02	18	643	27.95	0.51
AT-03	10	361	15.69	0.57
BR-01	16	373	16.21	0.36
BR-02	22	148	6.43	0.55
BR-03	30	98	4.26	0.45
BR-04	29	48	2.08	0.62
BR-05	28	38	1.65	0.52
BR-06	14	338	14.69	0.55
BR-07	10	237	10.30	0.48
BR-08	14	198	8.60	0.61
CN-01	35	97	4.21	0.63
CN-02	28	97	4.21	0.70
CN-03	22	97	4.21	0.55
DE-01	29	198	8.60	0.64
DE-02	18	381	16.56	0.46
ES-01	20	218	9.47	0.70
FR-01	11	676	29.39	0.52
FR-02	6	756	32.86	0.50
JP-01	21	97	4.21	0.69
JP-02	28	98	4.26	0.63
PT-01	5	277	12.04	0.56
TR-01	41	185	8.04	0.60
UK-01	10	118	5.13	0.62
UKN-01	25	98	4.26	0.51
UKN-02	30	98	4.26	0.64
UKN-03	20	96	4.17	0.59

In Figure 3.6(b), we have Videos X Interactions. It is possible to see that the video which had more interactions was BR-03, which was one of the most populated of the data set (i.e., 30 individuals), while the video which had less interactions was UK-01, which was one of the less populated (i.e., 10 individuals). It is interesting to note that the average interaction factor γ was higher in the UK-01 (i.e., 0.62) than in the BR-03 (i.e., 0.46), which suggests that the number of individuals actually affects the amount of interactions found, as we also observed in the simulations. Another important factor that affects the amount of interactions is the length of the videos. Long videos are more inclined to generate more interactions than short videos, simply because there is more time to interactions to occur. The CN-02 video sequence is a good example. It generated a small amount of interactions, even though it had a reasonable amount of individuals (i.e., 28) and a high average γ (i.e., 0.7).

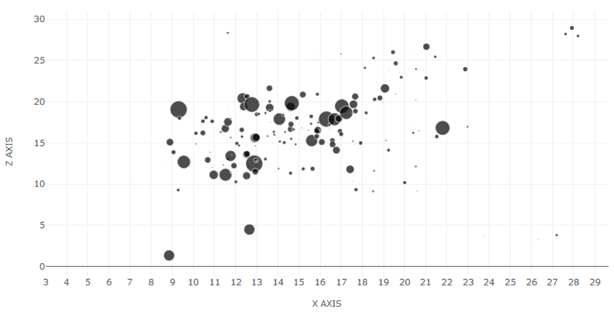


(a) Simulations X Interactions. (b) Videos X Interactions.

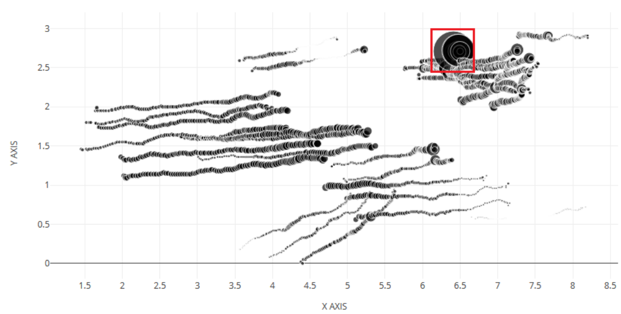
Figure 3.6: Total amount of interactions. In (a), we have Simulations X Interactions. The simulation that generated more interactions was the only one with Pink personality agents (100PinkPerson), while the simulation that generated less interactions was the only one with Blue personality agents (10BluePerson). In (b), we have Video Sequences X Interactions. The video that had the most interactions was BR-03, which was one of the most populated in the data set (that is, 30 people), while the video that had less interactions was UK-01, which was one of the less populated (i.e., 10 individuals).

The second visualization is a scatter plot, where we show the relationship between the interactions and the positions of the agents/individuals in the simulation/video, which can be seen in Figure 3.7. The X-axis represents the X-position and the Y-axis represents the Z-position. The size of the bubble represents the normalized amount of interactions in that position. In Figure 3.7(a), we have interactions in the 2D space (interactions x positions) in the 100HalfPinkHalfBlue simulation. Here, we expected to know in which parts of the environment more or less interactions occurred. Initially, we expected it to be random. Although, when visualizing the results achieved, we realized that simulations containing Pink agents seemed to have more dispersed interactions than simulations with Blue agents. It means, the visualization suggests that, when agents had a high interaction factor, they interacted almost anywhere in the environments, while agents with a low interaction factor tended to interact in the central areas of the environment. We believe that such behavior emerged because agents with a Blue personality had to have more agents around in order to be able to interact, while agents with a Pink personality were able to easily interact, even when there was just one other agent around. In Figure 3.7(b), we have interactions in the 2D space (interactions x positions) in video sequence BR-03. First, it important to clarify the difference between such interaction in the video sequences and the simulations. In our simulations, agents approached each other and interacted while idle. In the videos, individuals are usually moving through the environment, and interact between themselves even while in movement. In Figure 3.7(b), we can see that the interactions marked in the visualization seem to form the trajectory of the individuals which interacted. In addition, the larger marker that can be seen around the position $(6.5;\ 2.5)$, and highlighted with a red square, represents the interactions with a person which is idle, probably waiting to cross the street.

The third visualization is a network graph, where we show the relationships of all agents/individuals in a given simulation/video and can be seen in Figure 3.8, for both. To develop this view, it was



(a) Interactions in the 2D space (interactions x positions) in simulation.



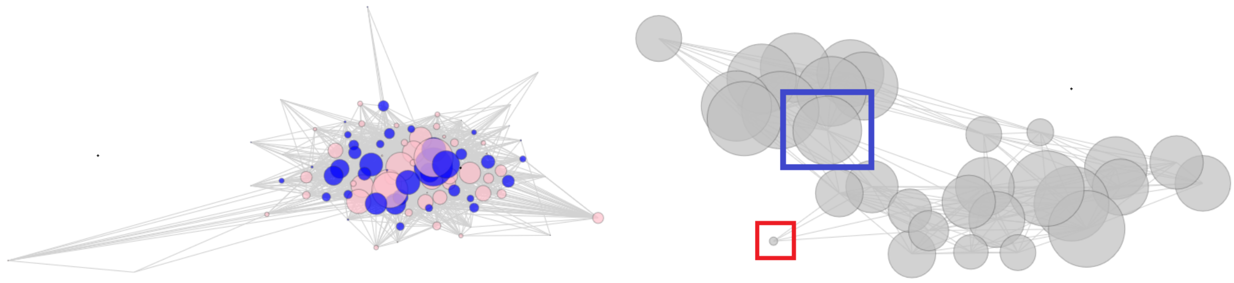
(b) Interactions in the 2D space (interactions x positions) in video.

Figure 3.7: Interactions in the 2D space (interactions x positions). In (a), we have the scatter plot of the amount of interactions by position in simulation 100HalfPinkHalfBlue. In the simulations, When agents had a high interaction factor, they interacted almost anywhere in the environments, while agents with a low interaction factor tended to interact in the central areas of the environment. In (b), we have the scatter plot of the amount of interactions per position in the BR-03 video sequence. The interactions marked in the visualization seem to form the trajectory of the individuals which interacted. The larger marker that can be seen around the position $(6.5; 2.5)$, and highlighted with a red square, represents the interactions with a person which is idle, probably waiting to cross the street.

necessary to use the *Networkx library*[7]. This library provides tools for developing graphs, vertexes, and edges. First step was to create an instance of an empty graph, then add vertexes and edges. The agents/individuals of each simulation/video are represented by the vertexes and the edges represent the relationships that each agent/person made. The size of each vertex is given by the number of interactions that the agent/person had, while the color of each node corresponds to each person used for simulations (i.e., Neutral, Blue and Pink). The idea of this visualization was to be able to easily see which interactive agents/individuals are related in the simulations/videos, that is, which agents/individuals have more relationship with others. In Figure 3.8(a), we have the interactions for simulation 100HalfPinkHalfBlue. As expected, agents with a Pink personality (pink nodes) are generally those who have had the most relationships and, therefore, are more likely to be in the center of the visualization, while agents with a Blue personality (blue nodes) are usually the most isolated. In Figure 3.8(b), we have the interactions for video sequence BR-03. Since in the video sequences we have varied OCEAN values (in the simulations, we had three fixed personalities), we assign no color to the nodes. As we already observed in the simulations, individuals which interacted with a lower number of other individuals are more isolated in the visualization than individuals which interacted with a higher number of other individuals. Although we have no colors to identify the personalities, when we hover the mouse over a node, a tool tip with informations about that person appears. With this, we were able to check the interaction factor of such individuals. One of the individuals who is most isolated and has a small node, highlighted by a red square in Figure 3.8(b), also has an interaction factor $\gamma = 0.22$, while the individuals represented by bigger nodes have interaction factors $\gamma \geq 0.5$ (for example, the person highlighted with the blue square has an interaction

[7]NetworkX is available at https://networkx.github.io/

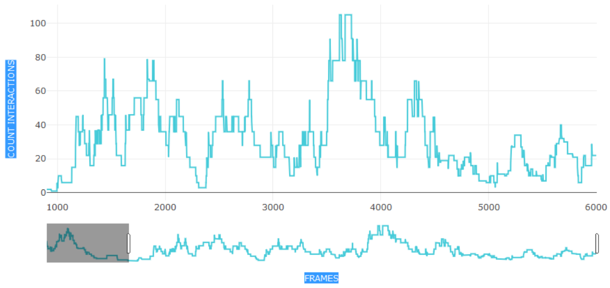
factor of 0.62). It seems to validate what was observed with the simulations, where low values of interaction factors (represented by the Blue personality) also generated less interactions than high values of interaction factors (represented by the Pink personality).



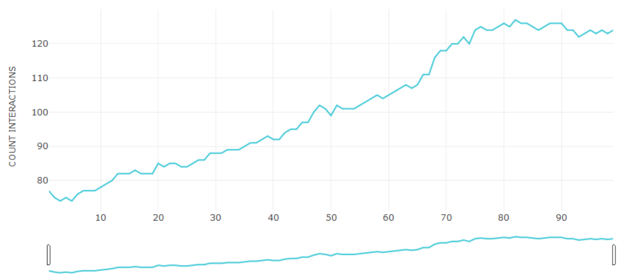
(a) Interactions between agents in simulation. (b) Interactions between agents in video.

Figure 3.8: Interactions between agents/individuals. In (a), agents with a Pink personality are generally those who have had the most relationships and, therefore, are more likely to be in the center of the visualization, while agents with a Blue personality are usually the most isolated. In (b), the person which is more isolated and have a small node (highlighted by a red square) has an interaction factor $\gamma = 0.22$, while some of the individuals represented by bigger nodes have interaction factors $\gamma \geq 0.5$ (as the person who is highlighted with the blue square, which has interaction factor of 0.62).

The fourth visualization shows a time-line chart and can be seen in Figure 3.9. It uses a slider to select an interval of frames to visualize. In such intervals, the amount of interactions for each frame is shown, for the chosen simulation. This visualization was proposed in order to be able to find if the interactions occur at a specific time of the simulations or videos. In Figure 3.9(a), we have the interactions by frame for simulation 100HalfPinkHalfBlue. As in the second visualization (i.e., scatter plot), we also expected that it would be something random. Although, when visualizing the results achieved, we perceived that the simulations with agents which had a Pink personality, presented more peaks of interactions in the initial/final frames than the simulations with Blue personality agents, which presented interactions more focused in the intermediate frames. We believe that such behavior can be explained by the way that agents interact in the simulations. As explained in Section 3.1.1, while agents are interacting, their respective interaction factors decrease. When they stop to interact, they can not interact again for a defined amount of time (i.e., 150 frames). So, we believe that agents with a Pink personality start to interact early in the simulation and, when such interactions finish, can just interact again 150 frames after. On the other hand, agents with a Blue personality took more time to interact among each other, since they had a low interaction factor. In Figure 3.9(b), we have the interactions by frame for video sequence BR-03. It is possible to notice that, for this video sequence, the amount of interactions increases as time passes. This behavior can be explained by the fact that people are more distant at the beginning of the video, approaching as they walk through the environment, as time passes by. The same occurs with the video AE-02, which is the second video in number of interactions (Figure 3.6(b)). On the other hand, the videos with a low amount of interactions had a different behavior: the amount of interactions kept varying through the time. It suggests that the quantity of individuals present in the video sequences (as well as their respective OCEAN inputs) affected the way they interacted in the video. In addition, the initial positioning of these people appeared to be relevant. Taking the video BR-03 as example: people are more distant at the beginning of the video, so the interactions increase when those individuals, in the video, approach and/or cross each other.



(a) Interactions by frame in simulation.



(b) Interactions by frame in video.

Figure 3.9: Interactions by frame. In (a), agents which had a Pink personality presented more peaks of interactions in the initial/final frames than the simulations with Blue personality agents, which presented interactions more focused in the intermediate frames. In (b), it is possible to notice that the amount of interactions grows up as the time passes by, for this video sequence (i.e. BR-03). The same does not occur for other video sequences (for example, UK-01), which suggests that the quantity of individuals present in the video sequences (as well as their respective OCEAN inputs) affected the way they interacted through the video, as well the initial positioning of individuals.

Final Comments Regarding Tested Visualization Techniques: The used visualization techniques and results were empirically evaluated in our research lab. Following are aspects that could be observed:

- The Bar chart was more adequate to understand the number of interactions at each video;
- The scatter plot was more adequate to understand the location of interactions;
- The network graph was more adequate to perceive the impact of personalities in the interaction events; and finally
- The time-line chart better provides visualization of evolution of interactions as a function of time.

Although we could understand group data, none of the studied techniques could show individual parameters in a understandable way. It is important to notice that we did not perform an exhaustive search for visualization methods in order to provide a way to visually understand internal parameters of agents. So, we chosen for developing a new viewer where we could create visual options specifically for parameters we are interested on. Our goal is to observe and analyse people perception as a function of a viewer that hypothetically is made for visualize internal parameters of individuals. Next section presents this viewer and the performed analyses.

4. PERCEPTION OF CULTURAL FEATURES IN CROWDS

In the previous section, we used data from people in real videos (taken from the *Cultural Crowds* data set) generated by the *GeoMind Software*[1] to obtain and visualize data regarding interpersonal interactions. In addition, we also simulate groups interactions in order to have ground truth. As stated before, none of the tested techniques could present in a visible and understandable way the internal parameters of agents in simulations, or pedestrians in real video sequences. The main goal of this section is to discuss and evaluate a viewer that has been developed to show geometric and non-geometric parameters of individuals. Geometric data are speed, distance, density and angular variation, while non-geometric data states for emotion and personality traits (originated by geometric data). This section is divided into two sections: *i)* Section 4.1 where the cultural features and questionnaires used to obtain people's perceptions are explained, and Section 4.2 where the results on perceptions are presented.

4.1 Methodology of Perceptions of Cultural Features in Crowds

This section is organized in three parts: *i)* Section 4.1.1 presents a brief explanation of how Favaretto et al. [21, 22, 24] calculated data on geometric and non-geometric features of people present in real videos; *ii)* Section 4.1.2 presents a viewer we develop that aims to represent the geometric and non-geometric features of people present in real videos; and finally *iii)* Section 4.1.3 presents the questionnaire on geometric and non-geometric (emotions and personality traits) features. The following sections detail these processes.

4.1.1 Explaining Data for Geometric and Non-Geometric Features

Based on the tracking input file, Favaretto et al. [21] compute following information for each pedestrian i, at each time step: *i) 2D* position $\vec{X}_i$ (meters); *ii)* speed s_i (meters/frame); *iii)* angular variation α_i (degrees) w.r.t. a reference vector $\vec{r} = (1, 0)$; *iv)* isolation level φ_i; *v)* socialization level ϑ_i; and *vi)* collectivity ϕ_i. To compute the collectivity affected in individual i from all n neighbors. Favaretto proposed $\phi_i = \sum_{j=0}^{n-1} \gamma e^{(-\beta \varpi(i,j)^2)}$, where the collectivity between two individuals (i, j) was calculated as a decay function of $\varpi(i, j) = s(s_i, s_j).w_1 + o(\alpha_i, \alpha_j).w_2$, considering s and o respectively the speed and orientation differences between two individuals i and j, and w_1 and w_2 are constants that should regulate the offset in meters and radians.

To compute the socialization level ϑ, Favaretto et al. [22] use an *Artificial Neural Network (ANN)* with a *Scaled Conjugate Gradient (SCG)* algorithm in the training process to calculate the socialization ϑ_i level for each individual i. The *ANN* has 3 inputs (collectivity ϕ_i of person i, mean

[1]GeoMind Software is available at https://www.rmfavaretto.pro.br/geomind/

Euclidean Distance from a person i to others $\bar{D_{i,j}}$ and the number of individuals in the *Social Space*[2] according to *Hall's Proxemics* [31] around the person i). The isolation level corresponds to its inverse, $\varphi_i = 1 - \vartheta_i$. For more details about that, please refer to [21, 22]. For each individual i in a video, it was computed the average of individual parameters generating a vector $\vec{V_i}$ of extracted data where $\vec{V_i} = \left[\bar{s}_i, \bar{\alpha}_i, \bar{\varphi}_i, \bar{\vartheta}_i, \bar{\phi}_i\right]$.

To detect the five dimensions of *OCEAN* for each person, Favaretto et al. [21] used the *NEO PI-R* [12] that is the standard questionnaire measure of the *Five Factor Model*. Firstly, they selected 25 from the 240 items from *NEO PI-R* inventory that had a direct relationship with crowd behavior. In order to answer the items with data coming from real video sequences, they proposed equations that could represent each one of the 25 items with features extracted from videos. For example, in order to represent the item "1 - Have clear goals, work to them in orderly way", Favaretto and other VHLab colleagues consider that the individual i should have a high speed s and low angular variation α to have answer in concordance with this item. So the equation for this item was $Q_1 = s_i + \frac{1}{\alpha_i}$. In this way, they empirically proposed equations for all the 25 items, as presented in [21].

In other work, Favaretto et al. [24] proposed a way to map *OCEAN* dimensions of each pedestrian in *OCC Emotion Model*, regarding four emotions: Anger, Fear, Happiness and Sadness. This mapping is described in Table 2.1, shown in Section 2. In Table 2.1, the plus/minus signals along each factor represent the positive/negative value of each one. For example concerning Openness, 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.

4.1.2 Developed Viewer

The viewer was developed using the *Unity3D*[3] engine, with $C\#$ programming language. The viewer allows users to rewind, accelerate and stop the video sequence through a time controller. Figure 4.1 shows the main window of the viewer, which 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;

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.

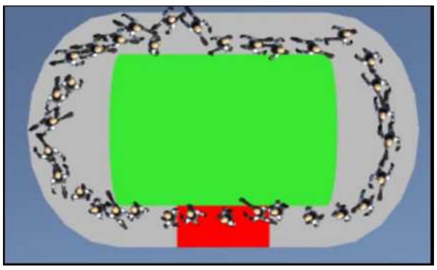
[2] *Social space* is related to 3.6 meters [31].
[3] *Unity3D* is available at https://unity3d.com/

3. Top-view cam image: a window that shows the top view of the environment, as if the user were looking at the crowd through the top;

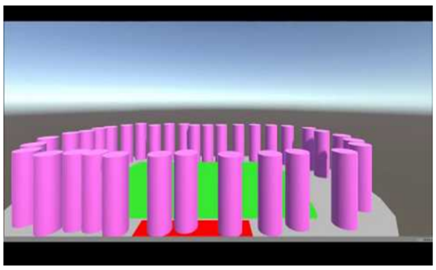
4. First-person cam image: area 4 shows first-person view from the viewpoint of a previously selected agent. This agent is highlighted in area 3;

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.

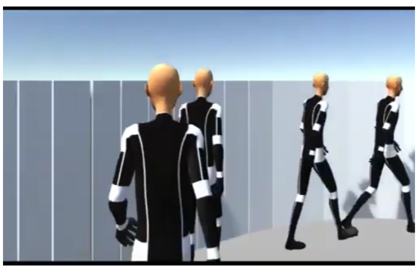
Figure 4.1: Main window of the viewer.



(a) Top view



(b) Oblique view



(c) First-person view

Figure 4.2: Types of visualization - (a) top view, (b) oblique and (c) first-person view.

This viewer has three modes of visualization: *(i)* first-person visualization, *(ii)* top view, and *(iii)* an oblique point of view. Figure 4.2 shows an example of each type of camera point of view in a video available in the *Cultural Crowds* data set. In addition to these different points of view, it is possible to observe all the pedestrians present at each frame f. Pedestrians can be represented by an humanoid or cylinder type avatar. Each pedestrian i present at frame f has a position (X_i, Y_i) (already converted from image coordinates to world coordinates). In addition to the positions, it is also possible to know if the pedestrian is walking, running or stopped at frame f, depending on the current speed s_i^f. If at frame f the current speed of agent i is greater than or equal to $\frac{0.08m}{f}$ which is equivalent to $\frac{2m}{s}$, considering $\frac{24f}{s}$, then the avatar is running. Values are defined based on the *Preferred Transition Speed PTS* [1]. The used values of the transitions can be seen in Equation 4.1, considering the current speed of the agent s_i.

$$Animation = \begin{cases} \textbf{Idle}, & \text{when } s_i == 0; \\ \textbf{Walk}, & \text{when } 0 < s_i < \frac{0.08m}{f}; \\ \textbf{Run}, & \text{when } s_i \geq \frac{0.08m}{f}. \end{cases} \tag{4.1}$$

Also, for the humanoid avatar type, each speed transition is accompanied by an animation transition, for example, if the current speed $s_i = 0$, then it does not change the animation (remaining stationary), but if its speed is $0 < s_i < \frac{0.08m}{f}$, then the animation changes for walking as well as if $s_i \geq \frac{0.08m}{f}$, the animation of the avatar changes to running.

The visualization of agent features is shown in the features panel, illustrated in Figure 4.1, 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 features in the agents. Figure 4.3 shows all possible icons that are related to the three options. For example, when the user selects the emotion status, 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 4.3(g-j).

Figure 4.4 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, its color is changed as the information fixed in the features panel, represented by an identifier (for example *Agent10*, who is highlighted in green). In addition to the agent identifier, there are the representative icons of its features: 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 4.4) which is running, is not a collective agent, is isolated and happy. All possible icons that can appear on the features panel are presented in Figure 4.3.

The viewer also provides a radial menu to show the features' values of a selected pedestrian. For this, when the user clicks on the identifier of a certain agent in the panel of features, the radial

(a) Walking

(b) Running

(c) Socialization

(d) Isolation

(e) Individualism

(f) Collectivism

(g) Anger

(h) Fear

(i) Happiness

(j) Sadness

Figure 4.3: Icons from the features of the viewer - possible icons shown in the features panel from a determined agent (all icons were taken from the internet in a simple google search). 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.

Figure 4.4: 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.

menu will appear[4]. Illustrated features in Figure 4.5 are presented in seven categories: *I - Speed*,

[4]The Radial menu template was taken for free from the website https://assetstore.unity.com/

II - Collectivity, *III - Interpersonal Distance*, *IV - Socialization and Isolation*, *V - Hofstede Cultural Dimensions*, *VI - Big-Five personality traits* and *VII - Emotions*. In the example, the Big-five personality traits (item V) of *Agent6* (highlighted in red) were represented in a graphical way, considering the max value of each dimension (OCEAN on the right of radial interface).

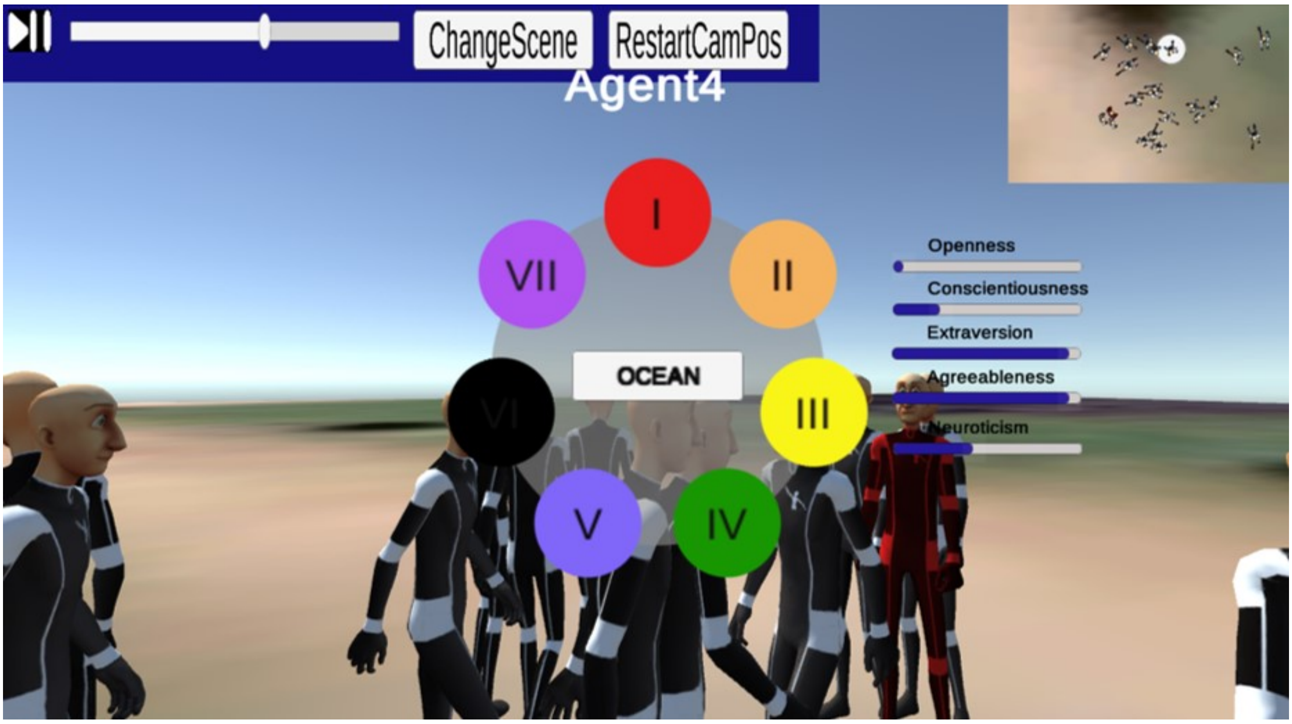


Figure 4.5: Radial menu of features: an example of the personalities shown in the radial menu from a selected pedestrian.

Figure 4.6 shows the visualization from a video recorded in Brazil with 15 pedestrians. In Figure 4.6(a), the oblique view is shown, where the user has a more general view of the experiment. In Figure 4.6(b), the first-person view is shown, where the user can feel as being part of the experiment. In both cases, the user can see a top view of the experiment in the upper right corner of the figures.

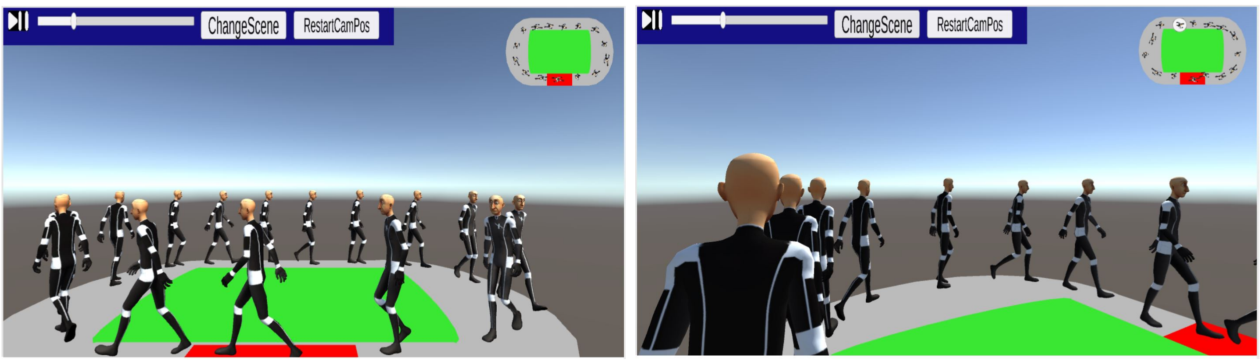


Figure 4.6: Visualization from a video of the FD experiment: different angle of visualizations: (a) oblique view and (b) first-person view.

Once we developed the viewer, we intend to test it with users. Next sections present details of this process.

4.1.3 Cultural Features Questionnaire

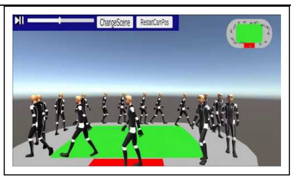
We formulated a questionnaire to evaluate how people perceive video/simulation features using our viewer. We divided it into two steps: *i)* The first aims to assess whether participants' perceptions of geometric features are affected by different viewpoints and types of avatars representation; and *ii)* the second aims to evaluate if participants can perceive emotions and personality traits. Before each question present on the survey, we present scenes of videos having virtual humans which positions came from the *Cultural Crowds* data set [20].

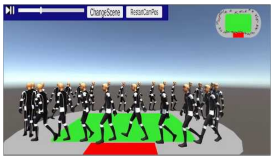
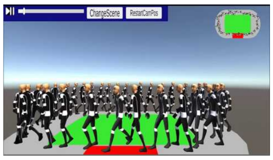
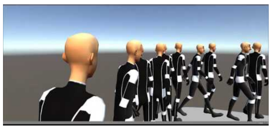
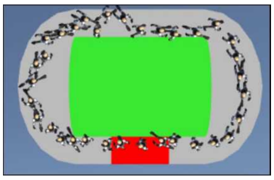
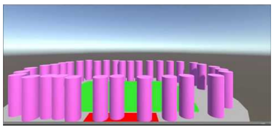
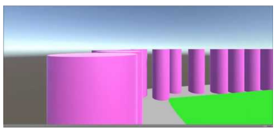
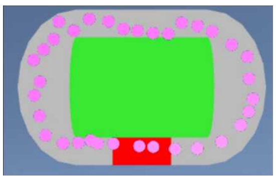
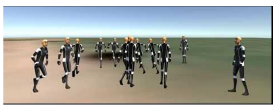
Table 4.1 shows some video characteristics that were used in this section, with information about the country where the video was recorded, the number of pedestrians and the density level (low, medium, or high). Each video was chosen based on the characteristics we want to verify in the questionnaire questions. For example, the videos used for speed questions were chosen based on pedestrian speed data. Thus, the videos $BR-15$, $BR-25$ and $BR-34$ were only used in scenes referring to questions of geometric features. The other videos were used in the scenes of questions of cultural features, however, the video $BR-01$ was also used in speed questions. All this data are visualized as shown in the previous sections, represented by cylinder or humanoid type avatars, which can be seen respectively in Figure 4.2(b) and (c). We also used three camera viewpoints (top view, oblique and First-person view illustrated in Figure 4.2(a), (b) and (c), respectively). Table 4.2 presents all the scenes used in the questions, the type of camera point of view, the type of avatar, and the *Cultural Crowds* video for each of them.

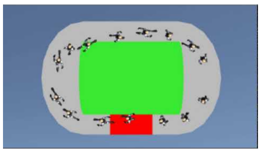
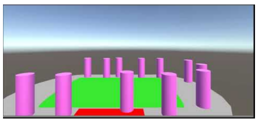
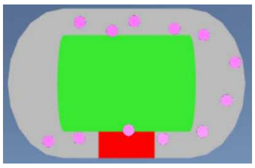
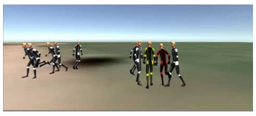
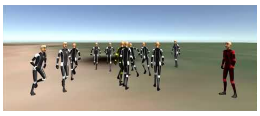
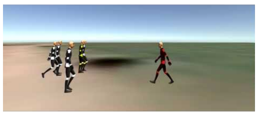
Table 4.1: Videos of the *Cultural Crowds* [20] data set.

Video	Country	N. Pedestrian	Density
AE-01	Unit. Arab Emirates	12	Low
AT-03	Austria	10	Low
BR-01	Brazil	16	Low
BR-15	Brazil	15	Low
BR-25	Brazil	25	Medium
BR-34	Brazil	34	High

Table 4.2: Scenes used in all questions in the survey. Walls were added in scenes 4* and 9* to evaluate density perception with and without walls. The last three scenes are related to the analysis of perceptions of emotions and personality traits. In addition, this table also shows the camera and avatar type of each scene, and the video identification (presented in Table 4.1) used.

Scene	Image	Camera	Avatar	Questions	Video
	Scene1	Oblique	Humanoid	D1, S2, S3, S5	BR-15

	Scene2	Oblique	Humanoid	D1	BR-25
	Scene3	Oblique	Humanoid	D1, D2, A1, A4, E1, E4	BR-34
	Scene4	First-Person	Humanoid	D2, D4, D6, A1, A5, E1, E4	BR-34
	Scene4*	First-Person	Humanoid	D5	BR-34
	Scene5	Top	Humanoid	D2, A1, A3, E1, E3	BR-34
	Scene6	Oblique	Cylinder	D3, A2, A4, E2, E4	BR-34
	Scene7	First-Person	Cylinder	D3, D4, A2, A5, E2, E5	BR-34
	Scene8	Top	Cylinder	D3, A2, A3, E2, E3	BR-34
	Scene9*	First-Person	Humanoid	D5	BR-15
	Scene10	Oblique	Humanoid	S1, S3	BR-01
	Scene11	Top	Humanoid	S1, S4	BR-01

	Scene12	Top	Humanoid	S2, S4, S6	BR-15
	Scene13	Oblique	Cylinder	S5	BR-15
	Scene14	Top	Cylinder	S6	BR-15
	Scene15	Oblique	Humanoid	Q1, Q2	AE-01
	Scene16	Oblique	Humanoid	Q3, Q4	BR-01
	Scene17	Oblique	Humanoid	Q5, Q6, Q7	AT-03

Part 1 Questionnaire - Geometric Perception

The first part of the questionnaire contains twenty-two questions, six related to the density, as shown in Table 4.3, with the correct answers highlighted in bold. In all density questions we asked **in which of the short sequences, presented to the participants, they observed the highest density level**.

The first question ($D1$) is a control question, i.e., we want to assess whether participants can perceive the density variation: low, medium and high density video scenes of individuals in crowds, respectively scenes 1, 2 and 3 shown in Table 4.2. Questions $D2$ and $D3$ aim to assess whether different points of view (camera types) influence the participants' perception of density. Our objective is to assess if participants can perceive the same density or if the perception of density changes due to the camera's point of view or the way agents are displayed. Before the two questions are presented scenes with the same density, but displayed with different points of view, where in $D2$ humanoid are used (scenes 3, 4 and 5) and in $D3$ are used cylinders (scenes 6, 7 and 8). Before question $D4$, two scenes (scenes 4 and 7) are presented with the same density and same point of view, but changing the type of avatar. This question aims to evaluate if different types of avatars influence the perception of density. In questions $D5$ and $D6$, walls were added around scenes 4 and 9 (see Figure 4.2(c)). Question $D5$ has the same objective as $D1$, to assess whether participants perceive the different types of density (low and high) in two scenes (4 and 9). Question $D6$ aims to assess whether walls influence density perception using the first-person camera. Before this question, scene 4 is presented twice, once with a wall and once without.

Table 4.3: Density questions and possible answers. The correct answer in highlighted in bold. $D1$ and $D5$ were made in order to be sure if the participants would know the concept of density. In $D5$ and $D6$ walls were added around the environment to evaluate density perception with and without walls.

Question	Possible answers
D1: In which scene do you see the highest density?	*a*) Scene 1; *b*) Scene 2; *c*) **Scene 3**; *d*) No difference; *e*) I don't know.
D2: In which scene do you see the highest density?	*a*) Scene 3; *b*) Scene 4; *c*) Scene 5; *d*) **No difference**; *e*) I don't know.
D3: In which scene do you see the highest density?	*a*) Scene 6; *b*) Scene 7; *c*) Scene 8; *d*) **No difference**; *e*) I don't know.
D4: In which scene do you see the highest density?	*a*) Scene 4; *b*) Scene 7; *d*) **No difference**; *e*) I don't know.
D5: In which scene do you see the highest density?	*a*) Scene 9 (walls); *b*) **Scene 4 (walls)**; *c*) No difference; *e*) I don't know.
D6: In which scene do you see the highest density?	*a*) Scene 4; *b*) Scene 4 (walls); *c*) **No difference**; *e*) I don't know.

Regarding speed perception, the questionnaire also contains six questions, as shown in Table 4.4, where all of them are related to low-density videos described in Table 4.1. The goal of these questions is to evaluate the speed levels running and walking, as presented in Equation 4.1, related to point of view (top and oblique cameras) and the two types of avatars: cylinder and humanoid. By having a lot of variation in the avatar's view, the camera in the first person is not evaluated on questions related to speed. Questions $S1 - S4$ are intended to assess whether perceptions of speed levels are influenced by two types of viewpoints (top and oblique camera). Question $S1$ presents two scenes (10 and 11, shown in Table 4.2) with “run" speed, and respectively oblique and top cameras. Same process for question $S2$ but with the “walk" speed, using scenes 1 and 12. Questions $S3$ and $S4$ present different speeds, respectively, in oblique (scenes 1 and 10) and top (scenes 11 and 12) cameras. With this, we can know if the speed perception is clearer in one of these two viewpoints. Finally questions $S5$ and $S6$ present two scenes containing two types of avatars, with oblique (scenes 1 and 13) and top (scenes 12 and 14) cameras, respectively, using the "walk" speed. These questions are intended to assess whether speed perceptions are influenced by the type of avatar.

Table 4.4: Speed questions and possible answers. The correct answer in highlighted in bold.

Question	Possible answers
S1: In which video did you observe the higher speed?	*a*) Scene 10; *b*) Scene 11; *c*) **No difference**; *e*) I don't know.
S2: In which video did you observe the higher speed?	*a*) Scene 1; *b*) Scene 12; *c*) **No difference**; *e*) I don't know.
S3: In which video did you observe the higher speed?	*a*) **Scene 10**; *b*) Scene 1; *c*) No difference; *e*) I don't know.
S4: In which video did you observe the higher speed?	*a*) **Scene 11**; *b*) Scene 12; *c*) No difference; *e*) I don't know.
S5: In which video did you observe the higher speed?	*a*) Scene 1; *b*) Scene 13; *c*) **No difference**; *e*) I don't know.
S6: In which video did you observe the higher speed?	*a*) Scene 12; *b*) Scene 14; *c*) **No difference**; *e*) I don't know.

Regarding angular variation, the questionnaire contains five questions with comparisons between the three types of cameras and two types of avatars. All angular variation questions use scenes from $BR-34$ video (high density), as shown in Table 4.1 and Table 4.2. Questions $A1$ and $A2$ aim to assess whether points of view influence the angular variation perception. $A1$ presents three scenes (scenes 3, 4 and 5, shown in Table 4.2) with humanoids viewed through three types of cameras. Similar process for question $A2$ (scenes 6, 7 and 8) where avatars are cylinders. Finally, questions $A3$, $A4$, and $A5$ present, respectively, two scenes containing the two different avatars with the top (scenes 5 and 8), oblique (scenes 3 and 6), and first-person (scenes 4 and 7) cameras. These questions aim to assess whether the types of avatars influence perceptions.

Regarding distance, the questionnaire also contains five questions (as shown in Table 4.6), all with videos containing high density. The distribution of the questions is the same as the questions on angular variation, using the same videos and scenes (shown in Table 4.2) in the respective

Table 4.5: Angular variation questions and possible answers. The correct answer in highlighted in bold.

Question	Possible answers
A1: In which scene do you observe more angular variation performed by the agents?	*a*) Scene 3; *b*) Scene 4; *c*) Scene 5; *d*) **No difference**; *e*) I don't know.
A2: In which scene do you observe more angular variation performed by the agents?	*a*) Scene 6; *b*) Scene 7; *c*) Scene 8; *d*) **No difference**; *e*) I don't know.
A3: In which scene do you observe more angular variation performed by the agents?	*a*) Scene 5; *b*) Scene 8; *c*) **No difference**; *d*) I don't know.
A4: In which scene do you observe more angular variation performed by the agents?	*a*) Scene 3; *b*) Scene 6; *c*) **No difference**; *d*) I don't know.
A5: In which scene do you observe more angular variation performed by the agents?	*a*) Scene 4; *b*) Scene 7; *c*) **No difference**; *d*) I don't know.

questions. Being the first two questions ($E1$ using scenes 3, 4 and 5 with humanoid, and $E2$ using scenes 6, 7 and 8 with cylinder) aiming to evaluate the perception regarding the points of view, and the last three questions (scenes 5 and 8 in $E3$, 3 and 6 in $E4$, and scenes 4 and 7 in $E5$) regarding the types of avatars.

Part 2 Questionnaire - Perception of Emotions and Personality Traits

The second stage of the questionnaire contains seven questions related to emotions and personality traits. In all scenes presented in the questions, two avatars of different colors (red and yellow) are highlighted to be the focus of the questions, as illustrated in Figure 4.7. All questions related to cultural features with their answers (correct answers are highlighted in bold) are presented in Table 4.7. We use as ground truth the results obtained by the approach proposed by Favaretto et al. [24]. In the example of Figure 4.7, the initial and final frames of $scene15$ (shown in Table 4.2) are shown, where there is a group of pedestrians (represented by avatars) in the right part of the scene. The avatar highlighted in yellow is part of this group and the avatar highlighted in red walk trough the group with a higher speed.

$Scene15$ is related to questions $Q1$ and $Q2$, where it is asked which avatar (yellow or red) is, respectively, neurotic and angry. Questions $Q3$ and $Q4$, related to $scene16$, are intended

Table 4.6: Distance questions and possible answers. The correct answer in highlighted in bold.

Question	Possible answers
E1: In which scene do you observe the largest distance among agents?	*a*) Scene 3; *b*) Scene 4; *c*) Scene 5; *d*) **No difference**; *e*) I don't know.
E2: In which scene do you observe the largest distance among agents?	*a*) Scene 6; *b*) Scene 7; *c*) Scene 8; *d*) **No difference**; *e*) I don't know.
E3: In which scene do you observe the largest distance among agents?	*a*) Scene 5; *b*) Scene 8; *c*) **No difference**; *d*) I don't know.
E4: In which scene do you observe the largest distance among agents?	*a*) Scene 3; *b*) Scene 6; *c*) **No difference**; *d*) I don't know.
E5: In which scene do you observe the largest distance among agents?	*a*) Scene 4; *b*) Scene 7; *c*) **No difference**; *d*) I don't know.

(a) Initial frame (b) Final frame

Figure 4.7: Initial (a) and final (b) frames from $Scene15$.

to assess whether participants perceive which highlighted avatar is open to experience and fear, respectively. This scene shows a yellow highlighted avatar interacting with a group of avatars and a red highlighted avatar standing alone with no interaction. Finally, questions $Q5$, $Q6$ and $Q7$ have the objectives of evaluating happiness, extroversion and sociability. Question $Q7$ was proposed after analyzing the results of question $Q6$, so it is explained in Section 4.2.6. These questions relate to $scene17$, which contains a yellow highlighted avatar walking with a group of avatars and a red highlighted avatar walking alone, in the opposite direction to all other avatars.

Table 4.7: Cultural questions and answers. Correct answers are highlighted in bold, according to Favaretto et al. [24].

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; *b*) **Red pedestrian**; *c*) Both pedestrians; *d*) Neither of them; *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; *b*) **Red pedestrian**; *c*) Both pedestrians; *d*) Neither of them; *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**; *b*) Red pedestrian; *c*) Both pedestrians; *d*) Neither of them; *e*) I don't know.
Q4: In your opinion, which of the two pedestrians highlighted in the video is afraid, yellow or red?	*a*) Yellow pedestrian; *b*) **Red pedestrian**; *c*) Both pedestrians; *d*) Neither of them; *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**; *b*) Red pedestrian; *c*) Both pedestrians; *d*) Neither of them; *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**; *b*) Red pedestrian; *c*) Both pedestrians; *d*) Neither of them; *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**; *b*) Red pedestrian; *c*) Both pedestrians; *d*) Neither of them; *e*) I don't know.

4.2 Results of Perceptions of Geometric and Non-Geometric Features in Crowds

This section aims to present the results of participants' perceptions regarding geometric data information (density, speed, angular variation and distance), personality traits and emotions.

Participants' responses are analyzed in order to answer whether perceptions about geometric features were influenced by different viewpoints and avatars, and if participants could perceive cultural features in avatars. The questionnaires were applied through social networks, i.e., all participants were volunteers. In case of boredom or tiredness, participants could stop answering the questions. No explanation of the research content was provided. Regarding the participants, an amount of 73 people volunteered for the experiment: 45 males (61.6%) and 28 (38.4%) females and 47.9% have some undergraduate degree. Regarding the age of the participants, 6.7% under 20 years old, 58.7% between 20 and 30 years old, 31.7% between 31 and 50 years old, and 4% over 50 years old. This section is organized in two parts: Sections 4.2.1, 4.2.2, 4.2.3, 4.2.4 and 4.2.5 discuss the results of perceptions about the geometric features of pedestrians, and Section 4.2.6 presents the results of perceptions about personalities and emotions.

4.2.1 Density Perception Analysis

With regard to density questions, Figure 4.8 shows the questions and the percentage of all answers, while Table 4.8 shows the percentage of correct answers for each question. In $D1$, 89% of participants responded according to ground truth, i.e., they were able to correctly classify the high density scene. In question $D2$, 70% chose one of the scenes, while 29% of the participants checked the option "I did not notice density difference". The details are shown in Figure 4.8, and the results indicates that the camera's point of view can disturb the density perception. As for the point of view, the oblique cameras presented the highest percentage of responses, being contrary to the work of Yang et al. [59], which people noticed greater densities through the camera in the first person. However, in this work, the authors did not use the oblique camera. In addition, the authors simulated virtual humans in more widely spaced positions, and simulations with higher densities. In question $D3$, 69% chosen one of the scenes, while 31% of the participants marked the option "I did not notice density difference", indicating that the visualization with cylinders or humanoids also change the final result. In question $D4$, 25% of people selected the option "I did not notice density difference", while 72% chosen one of the avatar types, being 41% of the participants have chosen humanoids. As with question $D1$, in question $D5$, most participants chose the highest density (85%) scene according to ground truth. The goal of these questions is to check whether walls alter the perception of density using the first-person camera. In this case ($D6$) 66% of participants answered that one of the scenes presented higher density in comparison to a same density scene without walls. In addition to this descriptive analysis, we performed statistical analysis using *Z-Test* to assess the significance of the right answer proportions of each question. Using the *P-Value*, Table 4.8, in the first cluster of answers regarding Density analysis, shows that almost all proportions questions were significant. Except questions $D1$ and $D5$, because they were control questions.

Table 4.8: All averages of right answer for each question. In addition, this table shows the *Z-Test* values used to assess the significance of these averages, with a significance level of less than 0.05.

Question	**Right Answer(%)**	**Z-Stats**	**P-value**
D1	89	22.12	< 0.01
D2	29	4.65	< 0.01
D3	31	5.20	< 0.01
D4	25	4.09	< 0.01
D5	85	19.66	< 0.01
D6	19	3.25	< 0.01
S1	32	5.20	< 0.01
S2	26	4.43	< 0.01
S3	48	7.51	< 0.01
S4	34	5.23	< 0.01
S5	17	3.05	< 0.01
S6	19	3.45	< 0.01
A1	14	2.67	< 0.01
A2	18	3.30	< 0.01
A3	24	3.85	< 0.01
A4	33	5.23	< 0.01
A5	28	4.50	< 0.01
E1	23	3.67	< 0.01
E2	24	4.29	< 0.01
E3	28	4.46	< 0.01
E4	27	4.24	< 0.01
E5	34	5.40	< 0.01
Q1	63	9.96	< 0.01
Q2	66	10.46	< 0.01
Q3	70	11.43	< 0.01
Q4	63	10.02	< 0.01
Q5	46	6.53	< 0.01
Q6	37	5.33	< 0.01
Q7	54	4.83	< 0.01

4.2.2 Speed Perception Analysis

In these videos, there was no analysis of perceptions using the camera in the first person, as we observed that these videos did not allow a good view of the scene. As shown in Figure 4.9, in $S1$, 32% of participants do not perceive any difference in speed while 64% chosen one of the scenes. In $S2$, 26% does not perceive difference while 74% chose one of the scenes (1 and 12). For questions $S3$ and $S4$, both in Figure 4.9 and in the percentage of right answers in Table 4.8, we can see that the oblique camera ($S3$) had a certain advantage, with 48% of participants answering the correct answer. Results were very similar in questions $S5$ and $S6$ having 17% and 19% respectively of people who do not perceive difference against 82% and 81% of people that chose one of the scenes. Therefore, our descriptive results indicate that camera point of view and type of avatar impacts

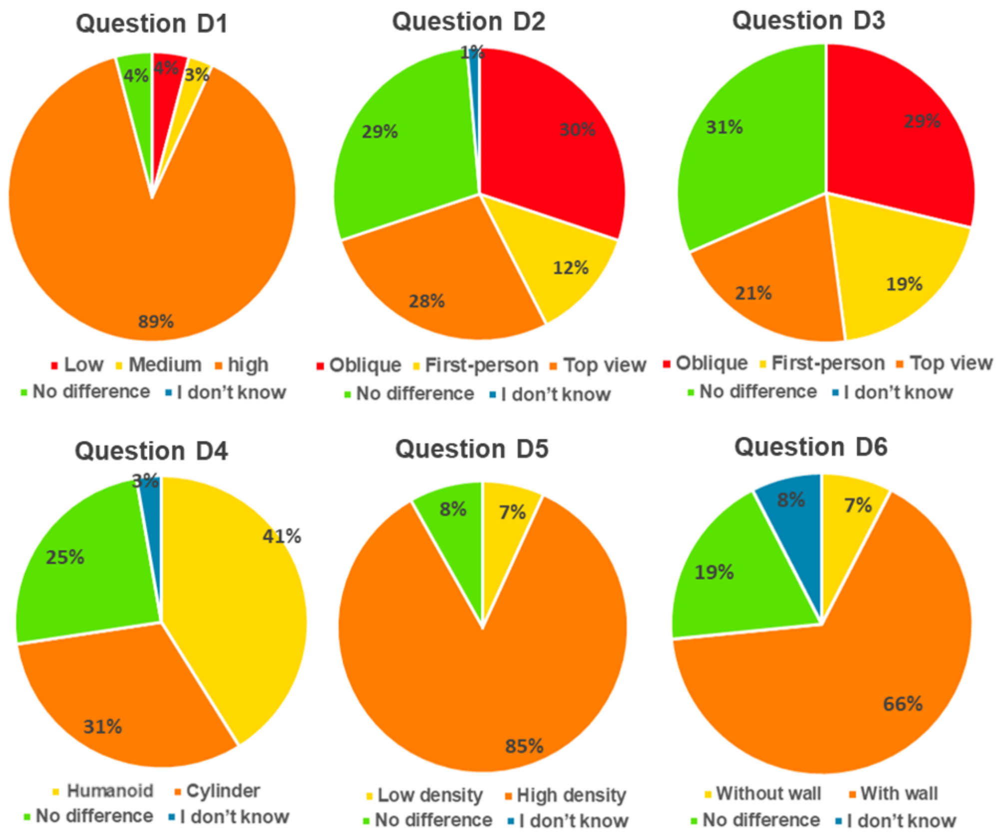


Figure 4.8: Perception concerning density - questions $D1$ to $D6$.

in the speed perception. In addition, considering the statistical analysis and using the *P-Value*, Table 4.8, in the second cluster of answers, shows that all proportions questions were significant.

4.2.3 Angular Variation Perception Analysis

In question $A1$, as shown in Figure 4.10, only 14% of participants do not perceive difference in the angular variation while 83% chosen one of the scenes and the top view camera was more selected. In $A2$, 18% of participants did not perceive difference while 79% selected one of the scenes. Most part of people who selected one video chose the one with humanoids. In questions $A3$, $A4$ and $A5$, most participants answered that one of the scenes had the largest angular variation of pedestrians, being 76% in $A3$, 66% in $A4$ and 69% in $A5$. So, our descriptive results indicate that the camera point of view and type of avatar impacts in the angular variation perception. In addition, considering the statistical analysis and using the *P-Value*, Table 4.8, in the third cluster of answers, shows that all proportions questions were significant.

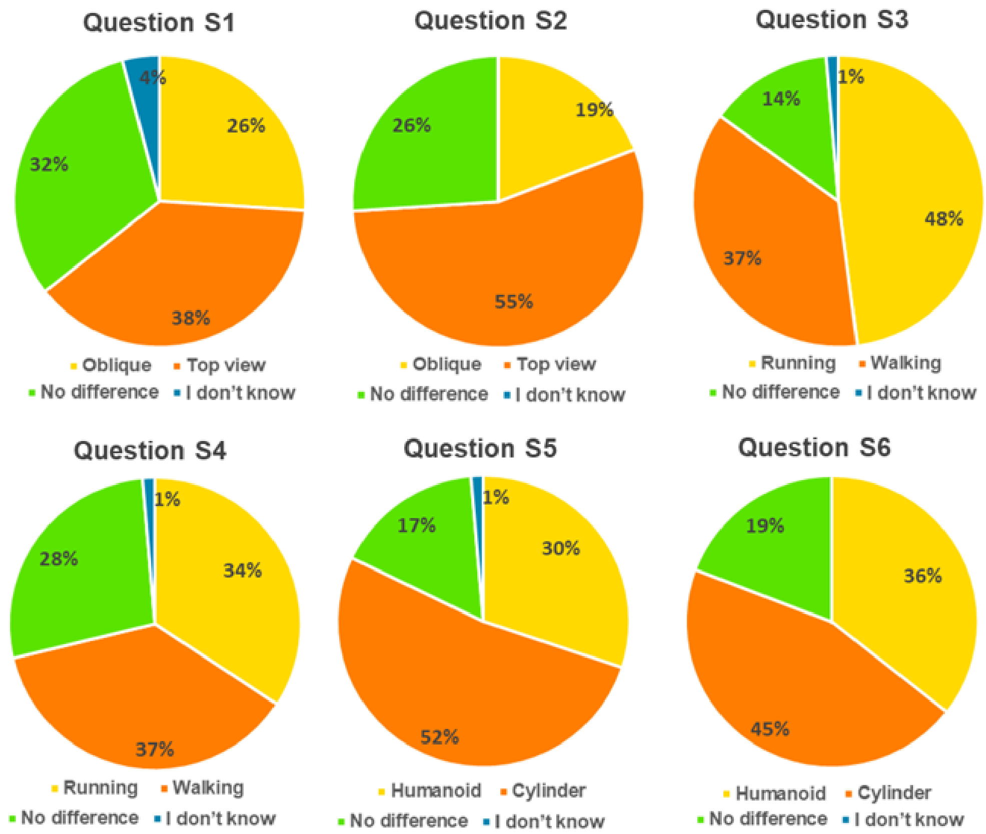


Figure 4.9: Perception concerning speed - questions $S1$ to $S6$.

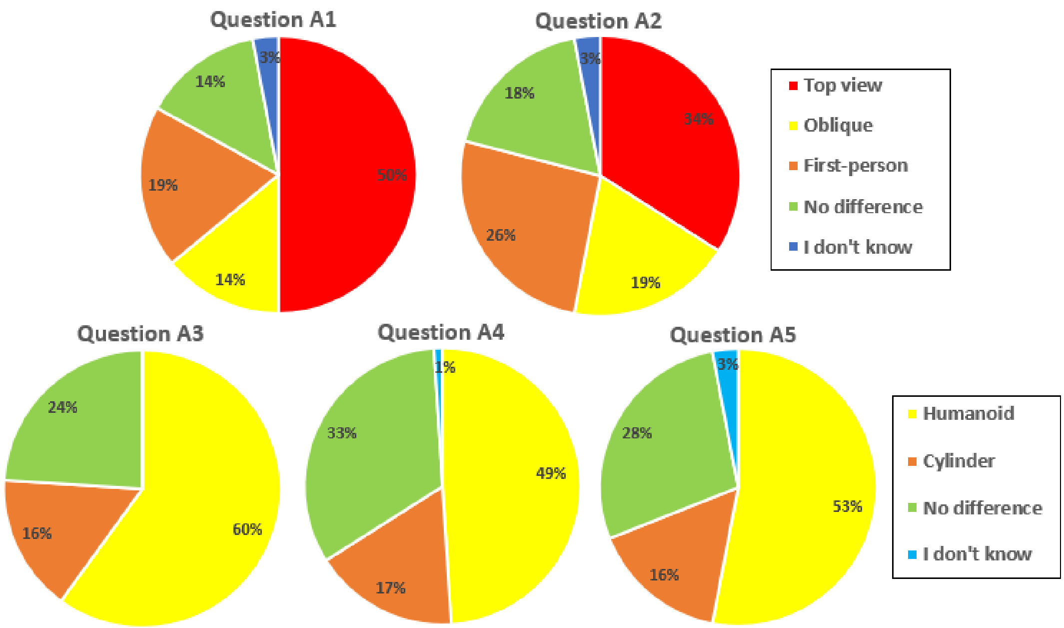


Figure 4.10: Perception concerning angular variation - questions $A1$ to $A5$.

4.2.4 Distance Perception Analysis

Indeed, results were very similar in both question $E1$ and question $E2$, as shown in Figure 4.11. In $E1$ we displayed humanoids with the three cameras (using scenes 3, 4 and 5, shown in

Table 4.2) and 22% of participants do not perceive differences, while in $E2$ we displayed cylinders (scenes 6, 7 and 8) and 24% also do not perceive changes. On the other hand, 77% and 73% of participants, respectively, selected one of the scenes in a approximately uniformly distributed way. Regarding the last three questions ($E3$, $E4$, and $E5$, which used, respectively, the same scenes as questions $A3$, $A4$ and $A5$), most participants answered that one of the scenes had the longest pedestrian distance, being 71% in $E3$, 73% in $E4$ and 66% in $E5$. So, our descriptive results indicate that the camera point of view and type of avatar impacts in the distance perception. In addition, considering the statistical analysis and using the *P-Value*, Table 4.8, in the fourth cluster of answers, shows that all proportions questions were significant.

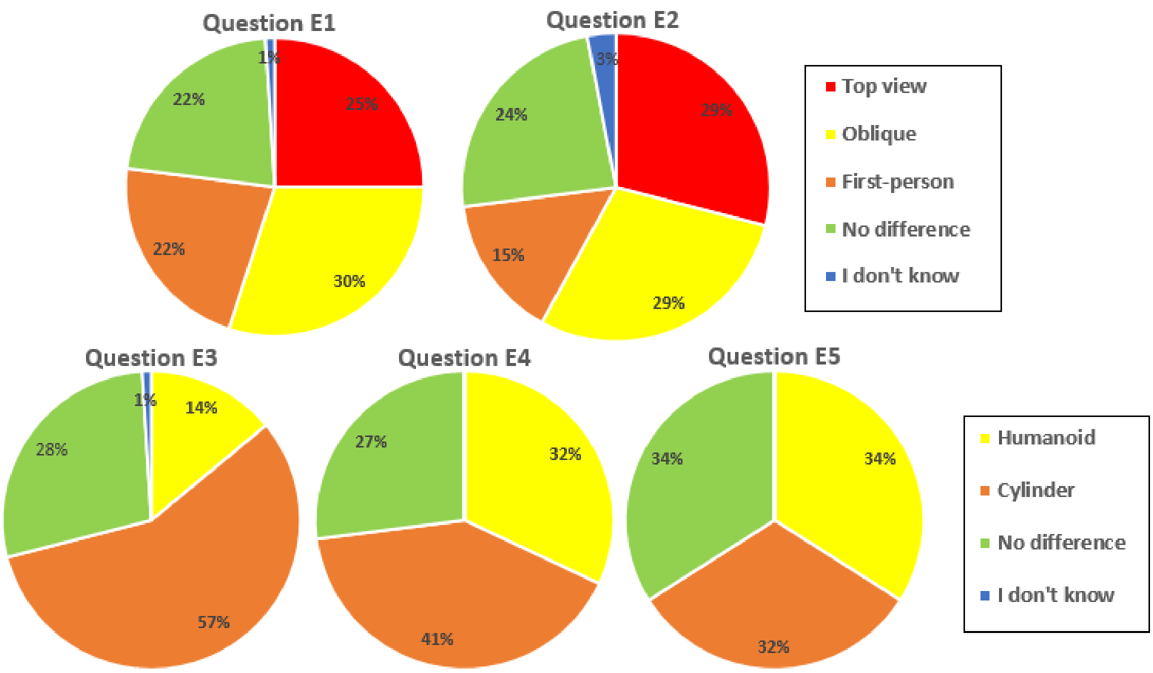


Figure 4.11: Perception concerning distance - questions $E1$ to $E5$.

4.2.5 Avatar and Camera Analysis

In the previous sections, analyzes were made in the contexts of each geometric feature and its questions. In this section, the focus is on the questions that analyzed the effect of avatars types (shown in Table 4.9) and camera types (Table 4.10) on participants' perceptions. In these analyzes we only considered the answers regarding the choices of avatars types (humanoid or cylinder) and camera types (first-person, top and oblique).

In Table 4.9, all questions are presented in which the focus was on the analysis of avatars types, the percentage of humanoid avatar choice, and the statistical analysis of this percentage through the *Z-Test* and the *P-Value*. Regarding the significance of the percentage of choice of avatars, we can see from the *P-Value* that all questions were significant. Overall the average choice of the humanoid avatar was 53.88%, where in 5 out of 9 questions, specifically $D4$, $A3$, $A4$, $A5$

and $E5$, over 50% of participants chose this type of avatar. To measure the significance of choice between one of the avatars, we performed a simple *ANOVA* analysis using a null hypothesis $H0$ that the average percentages of humanoid avatar choice would be equal to the cylinder representation. However, through the *P-Value*, shown in Table 4.11, we can see that the difference in the average choice percentages for one of the avatar types was not significant ($F_{1,16} = 0.65$, $p = 0.42$), not rejecting the null hypothesis. Thus, for the participants of this research, in all cases of perceptions of geometric features there was no difference in the choice of cylinder or humanoid. In another analysis, Table 4.11 shows that the average error percentages of questions focused on the analysis of avatars (considering all answers) types were much higher (73.88%), and significantly different ($F_{1,16} = 319.89$, $p < 0.05$), than the average of correct answers.

Table 4.9: All questions focused on the analysis of avatars types, and the percentage of humanoid avatar choice of each question. In addition, this table shows the *Z-Test* and *P-Value* used to assess the significance of each percentage, using the significance level less than 0.05.

Question	**Humanoid($\%$)**	**Z-stats**	**P-value**
D4	57	7.79	< 0.01
S5	37	5.09	< 0.01
S6	44	6.04	< 0.01
A3	79	13.69	< 0.01
A4	75	11.48	< 0.01
A5	77	12.31	< 0.01
E3	19	2.58	< 0.01
E4	44	5.77	< 0.01
E5	53	6.74	< 0.01

Table 4.10 shows the questions related to the analysis of camera types in the participants' perceptions, the percentage of choice of top camera in each of these questions, and the statistical analysis through the *Z-Test* to assess the significance level of each proportion. We can see from the *P-value* of each question that all proportions were significant. Overall, we can see from Table 4.11 that the average top camera choices (40.66%) are higher than the average of the other camera types (25% in first-person camera, and 34.33% in oblique camera). Through the null hypothesis $H0$ that the means of choice between the three cameras are equal, Table 4.11 shows that there was significance between these means ($F_{2,15} = 4.49$, $p = 0.02$), i.e., rejecting $H0$. However, for this analysis the questions $S1$ and $S2$ were taken, because there were no first-person camera analyzes. To make the analysis with all questions, we add the proportions of choice of cameras with similar angle of view (oblique and first-person). Thus, we propose the null hypothesis $H0$ that the averages of this sum of proportions and the top camera were equal. This hypothesis was not rejected because the comparison between the means was not significant ($F_{1,14} = 0.51$, $p = 0.48$). In addition, Table 4.11 also shows that the average error of these questions (considering all answers) was much higher (75.37%), and significantly different ($F_{1,14} = 261.4$, $p < 0.05$), than the average of correct answers.

Table 4.10: All questions focused on the analysis of camera types, and the percentage of top camera choice of each question. In addition, this table shows the *Z-Test* and *P-Value* used to assess the significance of each percentage, using the significance level less than 0.05.

Question	Top($\%$)	Z-stats	P-value
D2	40	5.20	< 0.01
D3	30	3.85	< 0.01
S1	59	7.62	< 0.01
S2	74	11.58	< 0.01
A1	60	8.89	< 0.01
A2	43	5.85	< 0.01
E1	33	4.53	< 0.01
E2	39	5.15	< 0.01

So, in this section we analyzed the participants' perceptions related to density, speed, angular variation and distances among agents displayed using two types of avatars and in three different cameras point of view. Through the significant difference in averages between the percentage of right and wrong questions in both camera and avatars analysis, the results in Table 4.11 indicate that human perception is affected by camera point of view and avatar representation. Regarding the choice of avatar type, there was no significant difference between the means of choice of avatar types, i.e., affecting equally the human perception. Regarding the choice of camera type, excluding speed questions, there was a significant difference between the means of choice, with advantage for the top camera, i.e., human perception is more affected by this type of camera.

Table 4.11: All analysis using simple *ANOVA*.

Dependent Variable	Independent Variables	Average	F	P-value
1. Answer Proportions with avatar questions (%)	Right answer Wrong answer	26.11% 73.88%	319.89	< 0.01
1.1. Avatar choice (%)	Humanoid Cylinder	53.88% 46.11%	0.65	0.42
2. Answer Proportions with camera questions (%)	Right answer Wrong answer	24.62% 75.37%	261.40	< 0.01
2.1. Camera choice without speed questions (%)	Oblique First.P Top	34.33% 25.00% 40.66%	4.49	0.02
2.2. Camera choice with speed questions (%)	Oblique + First.P Top	52.75% 47.25%	0.51	0.48
3. Answer Proportions with cultural features (%)	Right answer Wrong answer	57.00% 43.00%	4.83	0.04
3.1. Answer Proportions with cultural features (%)	Right + Partial right answer Wrong answer	64.35% 35.64%	32.45	< 0.01

Summarizing Findings Regarding Perception of Geometric Information from Crowds

Following aspects can be summarized based on last sections analyses:

- Regarding density issues, the results indicated that the perception of density is affected by the point of view and also by the avatar. With the addition of walls in scenes 4 and 9, we also came to the conclusion that the walls impact the perception of density;
- Regarding questions about speed, the results indicated that the perception of speed is also impacted by different points of view and different types of avatar. However, the oblique camera has a small advantage compared to the top camera;
- Regarding the questions about angular variation, the results indicated that the perception of geometric features is also impacted by different points of view and different types of avatar;
- With respect to distance questions, the results also indicated that the perception of distance is impacted both by different points of view and by different types of avatar;
- With regard to the general context about the types of avatar, we proved that the percentage of error in the choice of one of the avatars was higher and significantly different in relation to the percentage of correct answers, that is, the perceptions about the geometric features were impacted by the avatars;
- With regard to the general context about the types of points of view, we also proved that the percentage of error in choosing one of the cameras was higher and significantly different with respect to the percentage of correct answers. As a result, perceptions about geometric features were also impacted by different types of cameras.

4.2.6 Personality and Emotion Perceptions

Regarding the analysis of personalities and emotions, Figure 4.12 presents the answers to all the questions ($Q1 - Q7$) given by the participants. For these analyzes, we do not consider "I don't know" answers, we add the percentages of correct and partial correct (“Both pedestrians") answers, and the percentage of wrong answers. In the first two questions, it was interesting to see that more than half of the participants (66.17% in $Q1$ and 68.65% in $Q2$) answered according to the ground truth. The pedestrian highlighted in red was the most neurotic and angry, according to Favaretto's approach. Only a few participants answered incorrectly (33.83% in $Q1$ and 31.35% in $Q2$). As proposed by [24], geometrically, a neurotic person remains isolated and few collective. So, participants who do think that no agent was neurotic were certainly thinking about the psychological point of view, while we are analyzing based on space relationship. In $scene15$, the pedestrian highlighted in red has these characteristics. The pedestrian highlighted in red is: angry, isolated, low angular variation, low speed, low socialization and low collectivity, according to Favaretto [24].

Regarding questions $Q3$ and $Q4$, the results presented in Figure 4.12 show that most participants chose one of the right answers when compared to ground truth, i.e., 70.31% of participants correctly chose the yellow pedestrian as the most opened to experiences in $Q3$, and 67.14% correctly

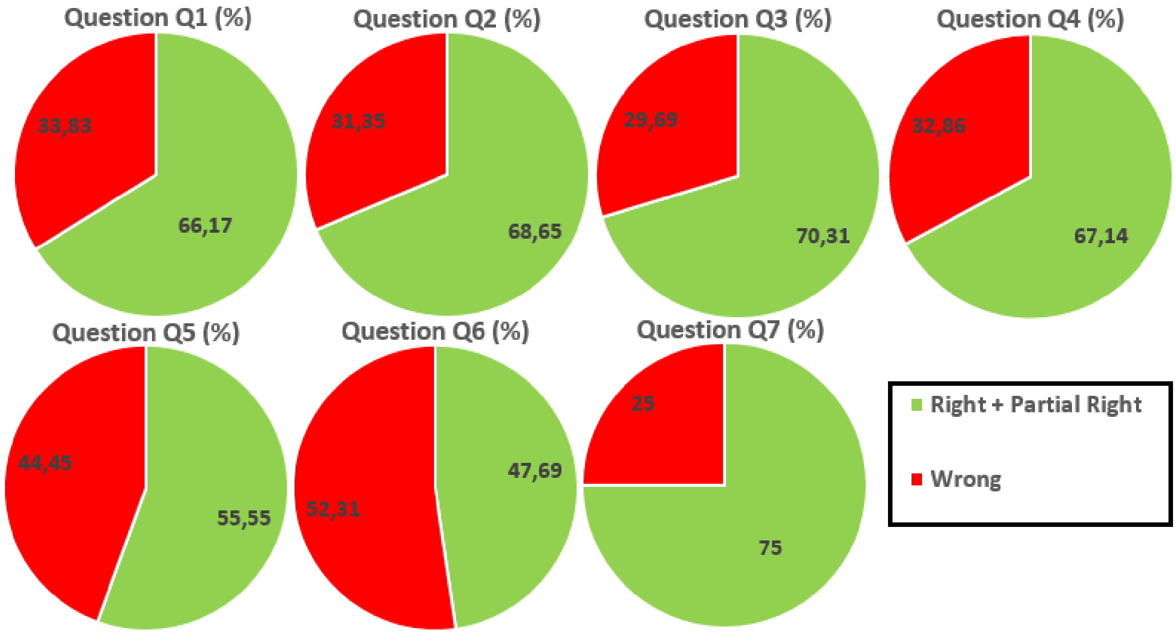


Figure 4.12: Perception analysis concerning Q1-Q7.

chose the red pedestrian as having fear in $Q4$. In the model of [24], a pedestrian opened to new experiences is related to a high value for the angular variation feature. Geometrically, according to what has been proposed in *Favaretto's 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.

Regarding question $Q5$, 55.55% of participants answered according to the ground truth. Geometrically, a happy person is not isolated and can present high levels of collectivity and socialization. Pedestrian highlighted in yellow presented that features and was correctly identified by the participants in the survey. Questions $Q6$ and $Q7$ analyze, respectively, extroversion and sociability. In question $Q6$, less than half of the participants (47.69% of them) answered correctly according to ground truth, indicating that the participants were not very sure about perceiving this feature. We believe that question $Q6$ 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 $Q7$, 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 $Q7$, most participants (75%) responded according to the model proposed by Favaretto et al. [24].

As shown in Table 4.8, all proportions of correct answers to these questions were significantly different from the percentages of wrong answers (without considering the right partial answers). In addition, as shown in Table 4.11, the average of correct answers of these questions was higher (57%), and significantly different ($F_{1,12} = 4.83$, $p = 0.04$), than the average of wrong answers. Considering the average of the sum of the partial correct answers with the right answers was much higher (64.35%) than the average of wrong answers. Thereby, Table 4.11 also shows that the sum average was significantly different ($F_{1,12} = 32.45$, $p < 0.05$) than the average of wrong answers. Thus, for this sample of people, the results of these two significant differences indicate that people can perceive cultural features in virtual humans without body and facial expressions.

Summarizing Findings Regarding Perception of Non-Geometric Information from Crowds

Following aspects can be summarized based on last section analyses:

- Regarding questions Q1-Q5, the results indicated that most participants were correct according to the ground truth. With that, we can see that in these questions people were able to identify cultural features (neuroticism, anger, opened to new experiences, fear and happiness) in virtual characters without facial and body expressions;

- Regarding question Q6, the results were not satisfactory. However, we believe that this is because the meaning of the word extroversion may caused confusion with the participants. With the addition of Q7 (sociable), the participants responded according to the ground truth. With that, we can also say that people were able to identify the cultural features in virtual characters without facial and body expressions.

5. PERCEPTION OF CG CHARACTERS

In the previous sections, we assessed perceptions in the context of pedestrians in crowds, while this section revisits the *Uncanny Valley* theory to observe its effect on people's perceptions of characters created with *CG*. This section is divided into two sections: Section 5.1, where it is explained about the characters created with *CG*, about the interaction with the *VR* environment, and about the questionnaire created to obtain people's perceptions, and Section 5.2, where the results about the perceptions are presented.

5.1 Methodology of Perception of *CG* Characters

This section is divided into three parts: Section 5.1.1 presents a brief description of Flach et al. [26], which also aimed to evaluate some *CG* characters, developed until the year 2012. In addition to the characters evaluated in [26], we included more recent characters (from the last 5 years); Section 5.1.2 describes the implemented *VR* system and finally; and Section 5.1.3 discusses the questionnaires applied. *Uncanny Valley* is represented by a *2D* chart, where the X-axis indicates the level of character realism from less to more realistic (from left to right, having higher values for realism on the right). The Y-axis defines the people's perception (in $\%$) regarding the comfort when watching/interacting with the characters (in our approach, we use the term comfort in our results). It goes from less to more comfortable where less comfortable is associated to small values in the Y-axis. We used only positive values in both axis.

5.1.1 The Characters

The first stage of this analysis was the selection of the characters. As mentioned before, we chose to reproduce the work of Flach et al. [26] because all used data was available to re-conduct the experiments. To reproduce the work of Flach et al., we use the same 10 characters analyzed by the authors. All of these characters are listed in the Figure 5.1 from (a) to (j). In addition, each character is accompanied by a legend of its origin, which may be a movie, game, or an animation found on the Internet from various origins. From (k) to (v) there are 12 characters created in last 5 years. In addition, the last three - from (w) to (y)- are the human models used in the *VR* environment.

For the most recent characters, from (k) to (v), we have tried to limit the choice of movies, games, series, among others, to a maximum of five years ago. With this, as proposed by Flach et al. [26], we evaluated the human likeness criterion. It contributes with the order the characters are placed in the horizontal axis of the *Uncanny Valley* chart, i.e., listing from left (small value - less realistic - in this axis) to right (higher values - more realistic). The vertical axis is responsible for, based on human perceptions, quantifying the uncanny feeling or comfort perceived by people. To ensure the characteristic of human likeness, we chose some characters that could represent a human being in a more realistic way, as shown in the cases (k, p, t and v) in Figure 5.1. These cases attempt to present simulated virtual humans with high levels of realism. Therefore, we need characters that escape from realism, counteracting the others cited earlier, such as the cartoon characters shown in Figure 5.1(m, n, q, s and u). This counterpoint is needed to form the horizontal axis of the *Uncanny Valley*.

(a) Internet (b) Internet (c) Internet (d) Movie - The Incredibles 1 (e) Obama's Cartoon (f) Internet (g) Movie - Cloudy with a Chance of Meatballs

(h) Movie - Beowulf (i) Game - Heavy Rain (j) Movie - Rango (k) Internet (l) Movie - Alita (m) Movie - How to train your dragon 2 (n) Movie - Thor Ragnarok

(o) Movie - Rogue One (p) Series - Love, Death and Robots (q) Movie - Moana (r) Game - Overkill's The Walking Dead (s) Movie - Spider-Man Into the Spider-Verse (t) Internet (u) Movie - The Incredibles 2

(v) Internet (w) Cartoon (x) Non-Realistic (y) Realistic

Figure 5.1: All the characters used in our approach. From a to j there are the characters used in the work of Flach et al. [26], from k to v are the most recent characters added in this analysis (all the characters' pictures have been taken from the internet), and finally from w to y we have the characters included in the interactive experiment.

In addition to the realism factor, we set some restrictions for the choices of the characters: *i)* the character has to represent a human being, that is, avoiding animals, for instance; *ii)* it should not be placed in an unreal place; *iii)* the character should wear normal (and not) minimal clothes to avoid distortions in perceptions; and finally *iv)* the scene should be focused on the character's face, so the participants could catch the movement of the mouth, the eyes, among other expressions. All of these restrictions were used to avoid possible negative influences on human perceptions regarding the ratings of the images and videos of the characters. Figure 5.1 shows all the characters used in this search.

5.1.2 Interaction with Virtual Characters

In order to evaluate whether human perception is impacted the more characters are included, we developed a *VR* application varying the following variables: realism of characters and number of characters. The application was developed using the *Unity3D* engine and the $C\#$ programming language. First of all, we asked to all participants to get close to the groups and observe them with the three types of characters and two levels of densities. Then, we applied a questionnaire firstly evaluating the images and videos of the characters (Figure 5.1 - w, x and y) by asking the participants to answer about their perception. In addition, participants tested the *VR* environment (6 times - one for each agent and density) and responded to the survey again regarding their perception as a function of the group presence and *VR* experience. The *VR* interaction was made using an *HP Mixed Reality Headset* [1] in a *3D* environment with three types of human models representing the characters and two density levels with these models: 0.26 and 0.65 agents per sqm. The size of the environment is 9 by 13 meters. In Figure 5.2, we show the virtual environment containing the character (x) of Figure 5.1 in the two different densities (a) and (b). Such characters models were chosen using the same realism criterion quoted in the previous section, and this was done to shape the horizontal axis of the *Uncanny Valley* chart, i.e., from the least realistic (character w), non-realistic (x) and realistic (y).

(a) Low Density　　(b) High Density

Figure 5.2: Virtual environment with model x (non-realistic 3D interactive model) presented in two types of density, low (10 characters) in (a), and high (25 characters) in (b).

[1] https://www8.hp.com/us/en/campaigns/mixedrealityheadset/overview.html

5.1.3 The Questionnaires

We formulated two surveys to conduct our research. First, as we wanted to compare the effects of *Uncanny Valley* on human perceptions about *CG* characters, we used the same structure of the questionnaire from Flach et al. [26], as presented in Table 5.1. Before answering the survey, participants received no explanation of the original intent of the research. This was done to avoid any type of influences on the participants' responses. All participants were asked if they agreed to cede their answers and personal information to our survey regarding age, gender, educational level and *CG* familiarity. The second questionnaire aimed to evaluate the interactive environment of *VR*. Next sections present the surveys.

Table 5.1: Questions regarding human perception applied to the participants.

Question	Possible answers
Q1: Do you think that the character in the picture/video above is:	*a*) A real person *b*) Created with CG *c*) Don't know
Q2: If created with *CG*, how realistic does it seem?	*a*) Very realistic *b*) Moderately realistic *c*) Unrealistic *d*) Don't know
Q3: Do you know this character?	*a*) Yes *b*) No *c*) Don't know
Q4: How do you would describe it?	*a*) Charismatic *b*) Non-Charismatic *c*) Don't know
Q5: Do you feel some discomfort (strangeness) looking to this character?	*a*) Yes *b*) No *c*) Don't know

First Questionnaire

This questionnaire was used to evaluate perception of all the characters in Figures 5.1, with the exception of the last three characters (w, x and y) because they have been analyzed in a second survey (as shown in the next section). The process is divided into two steps, the characters being shown randomly: In the first step an image of each character was shown before all the questions as referred in Table 5.1, while in the second step a video was shown for each character before all these questions as well. These steps compare the level of comfort people feel when observing characters in the pictures, where they are static, and in the videos where characters are moving. The level of comfort is asked in question Q5, shown in Table 5.1. Question Q1 introduces the character by asking whether it is created with *CG*, or is a real person, serving as control question. In question Q2, people were asked how realistic the character was. As shown in the work of Katsyri et al. [39], human likeness can be varied in an almost infinite number of different ways. Question Q2 covers such realism criterion and the responses influenced the order which the characters are placed in the horizontal axis (Human-Likeness) of the *Uncanny Valley* chart. We introduce the following equation:

$$\sigma_c = (NC\frac{\alpha_c}{10}) + (\frac{NC}{2}\frac{\beta_c}{10}) + (\frac{\gamma_c}{10}), \tag{5.1}$$

where σ_c is the realism factor of each character c. NC is the total number of characters (in our case $NC = 22$), α_c, β_c and γ_c are, respectively, the percentages of all "Very realistic", "Moderately realistic" and "Unrealistic" responses obtained in the applied survey, for character c. In addition, σ_c is normalized in interval $[1; NC]$ by the highest obtained realism factor among the NC characters: $\Omega_c = \frac{\sigma_c}{MAX_\sigma}NC$, where Ω_c is the normalized value and MAX_σ is the highest value of all the characters' realism factors. With this, the horizontal axis of the *Uncanny Valley* chart is shaped by the increasing order of the Ω value of NC character. If two characters have the same value for σ, we use also α, β and γ until a decision is taken.

The vertical axis (Comfort) is given by the percentage of the "No" answers of the Q5 question, that is, the more comfort the character presents, the higher is this value on the *Uncanny Valley* chart. In question Q3, people need to answer whether they know the character or not (familiarity). This question was asked to evaluate if familiarity with the evaluated character influences the comfort responses of Q5. Question Q4 was asked to assess whether the personality perception of the characters (whether it is charismatic or not) also influences the comfort responses of the Q5 question.

Second Questionnaire

The second questionnaire was related to only interactive characters viewed in Figure 5.1 (w, x and y). This questionnaire has been divided into three stages and aims to help answer the questions "What happens to the *Uncanny Valley* if, in addition to the images and videos, we include interactions with the evaluated characters?" and "How does the *Uncanny Valley* manifest if we have groups/crowds of characters and not only one agent?". The first two steps were identical to the first questionnaire explained in the previous section, but using images and videos of characters illustrated in Figure 5.1 (w, x and y). The third step involved the interaction of the participants with such characters in a *VR* application[2]. We created a room of 9x16 meters and changed the number of characters as well as their realism from Figure 5.1 (w) to (y). We asked the participants to visit the room and observe the agents, being as close as possible to them. At this stage, we tested 10 agents, with density of 0.26 agents$/m^2$, and 25 characters (0.65 agents$/m^2$) of the same type, i.e., firstly, character (w) from Figure 5.1, then the character (x) and finally character (y). Afterwards, participants answered the second survey where questions Q1, Q4 and Q5 from Table 5.1 were also

[2]I would like to thank Professor Dr Márcio Pinho from the GRV Lab and his undergraduate student Rafael Weiss, for developing the application and lending the *VR* equipment for this research.

considered. We did not evaluate Q2 because we chosen explicitly interactive agents in three known categories: low resolution, cartoon and realistic. The same applies to Q3 because all of them are Unity agents and they do not represent people who exist in real life. In addition, we propose one new question to be answered in order to evaluate the human perception in the *VR* application, as presented in Table 5.2.

Table 5.2: Question considered in addition to some of questions from Table 5.1 regarding human perception applied to the participants in *VR* environment.

Question	Possible answers
Q6: How do you feel about the amount of people around you:	*a*) Uncomfortable *b*) Comfortable *c*) Don't know

Once having prepared the surveys, we proceeded to perform the tests with participants. Section 5.2 detail this procedure.

5.2 Results of Perception of *CG* Characters

This section presents the results obtained from people's perceptions of characters created with *CG*. Section 5.2.1 presents the results obtained from the first questionnaire, which we intend to answer to the following questions: *i)* "Does the exposure to virtual characters, which has been going on for several decades now, reduce the Uncanny effect on people's perceptions?", and *ii)* "How the charisma and familiarity with virtual humans correlate to the *Uncanny Valley*?". Section 5.2.2 presents the results related to the second questionnaire, in order to answer the following questions: *iii)* "Does Interactive Environments impact in *Uncanny Valley* effect?", and *iv)* "How is our perception impacted if more than one character is presented instead of only one?".

5.2.1 Data obtained in Questionnaire 1

Results discussed in this section were obtained with the first questionnaire shown in Section 5.1, which was related to the characters in Figures 5.1, without (w, x and y). This questionnaire was applied on social networks, and all participants were volunteers. It was answered by 119 participants, where further data about them are shown in Table 5.3. In addition, all "I don't know" answers were discarded from the analysis.

Comparing Answers Obtained in 2012 and 2019

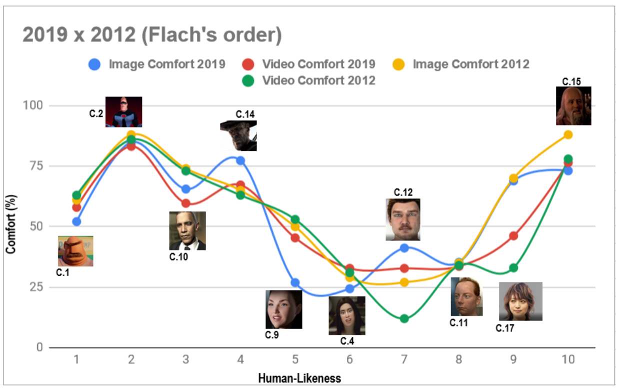
Firstly, it is important to mention that Flach et al.'s work [26] used another order for the characters' likeness (*X*-axis in *Uncanny Valley* chart), while we computed this using Equation 5.1.

Table 5.3: Information regarding the participants' (gender, age, educational level, and familiarity with *CG*).

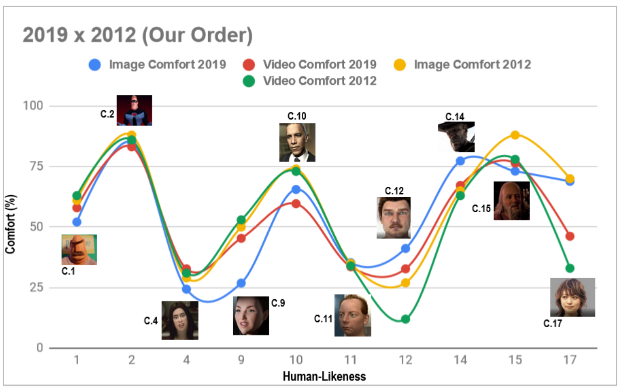
Genre	(%)	**Age**	(%)	**Educational Level**	(%)	**CG Familiarity**	(%)
Female	68.1%	0-20	40.3%	Incomplete High School	31.9%	Yes	68.07%
		20-30	37%	Complete High School	31.9%		
Male	58%	31-50	16.8%	Complete Graduation	21.8%	No	31.93%
		50+	5.9%	Complete Post-graduation	14.3%		

We compare such two orders for only the characters evaluated in 2012. Figure 5.3(a) shows the effects of *Uncanny Valley* on the results of the participants' perceptions (in 2012 and 2019) using Flach et al.'s human likeness (*X*-axis order). On the other hand, Figure 5.3(b) shows the same participants' perceptions using our human likeness order, for the same characters. The average value of comfort in the image analysis of Flach's work is 58.70% (standard deviation$=22.72\%$), and the video is 52.60% (standard deviation$=24.12\%$). In the present case, the evaluation with same characters obtained 54.95% (std$=21.95\%$) for image comfort, while 53.52% (std$=18.34\%$) for video comfort. We propose the null hypothesis $H0$ defining that, over the course of seven years, such *CG* characters are equally comfortable regarding human perception.

To evaluate this hypothesis, we performed a *Chi-squared test* with a significance level of 5% to evaluate the comfort of people when observing characters from Flach's work in 2012 and nowadays - 2019. We obtained respectively for image and video the following *p-values*: $1.16E-4$ and $6.85E-7$, so rejecting $H0$, i.e., obtained values of comfort are significantly different when compared perceptions in 2012 and 2019. Consequently, **people in 2012 were more comfortable with such old characters than in 2019**. A possible explanation is that this may happen due to the fact that nowadays we are more exposed to graphics of better quality.



(a) Flach order



(b) Our order

Figure 5.3: All the characters used in the work of Flach et al. [26] with Flach's order in (a), and our order in (b).

In another analysis, we computed the average value of comfort when analyzing the 22 characters tested in present work. While 10 of them are shown in Figure 5.3, and the new characters are presented in Figure 5.4. The average comfort of all characters is 69.13% (std$=22.74\%$) in image

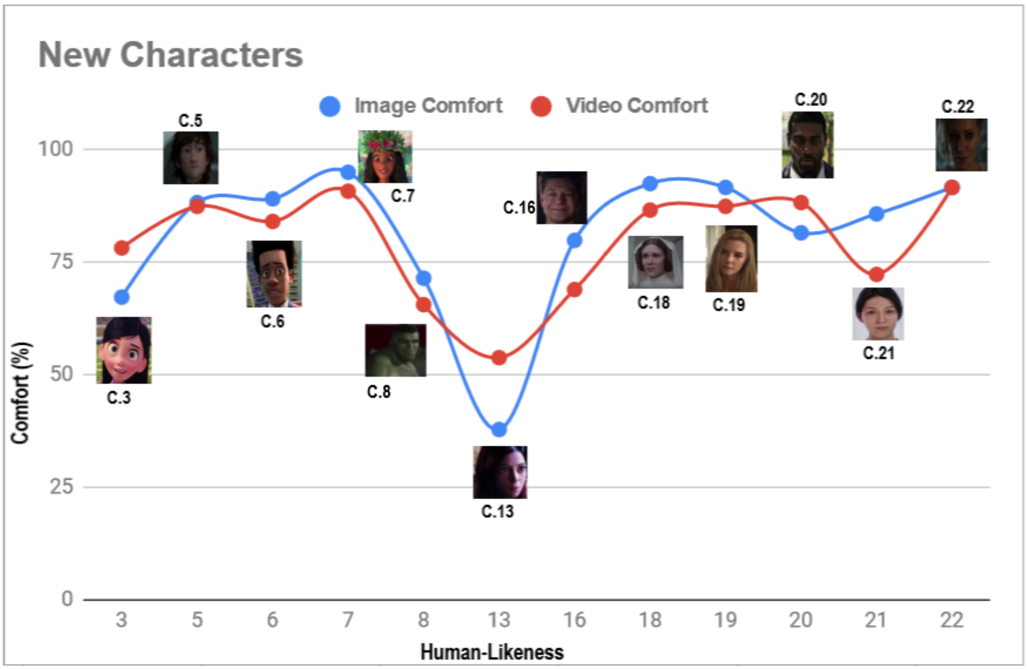


Figure 5.4: Recent characters evaluated in our approach.

and 67.72%(std=19.86%) in video analysis (see Figure 5.5). It presents higher average values than in Flach's work (58.7% with standard deviation of 22.72%, and 52.60% with standard deviation of 24.12%), respectively for image and video. We propose a null hypothesis $H0$ defining that the comfort in 2012 and 2019 are equal. We performed *T-test* and obtained the *P-values*$=0.12$ and 0.06 (respectively for image and video) with two different number of samples (10 characters in 2012 and 22 characters in 2019). Considering 30 degrees of freedom and a significance level of 5%, in the probability table, we found value$=2.089$, so rejecting $H0$ for both cases (image and video). It indicates that the comfort in those two analysis (2012 and 2019) are significantly different, i.e., **people in 2019 present higher comfort while evaluating virtual characters than in 2012**. It is important to mention that we tried to include characters varied in the same way as the sample presented in 2012, i.e., having cartoons and very realistic characters as well.

Regarding specifically the analysis performed in the present work (Figure 5.5), the level of comfort of each character was similar in the images and in the videos, except for some characters such as C.9, C.13, C.17 and C.21 in Figure 5.5. Just in some of them (C.17 and C.21) the video presented less comfort, as proposed in the *Uncanny Valley* theory [49]. In addition, character 17 was also used in the work of Flach et al. [26] and its result was similar to ours (approximately 46% of comfort in video and 68% in image).

Therefore, it is interesting to see that on the right side of Figure 5.5 we have the characters that are more realistic, while on the left we have the cartoon and not realistic ones. Another aspect can be considered regarding the realism, where character 4 (from the game Heavy Rain) was the 5th (from 10) most realistic character in Flach's work, and Beowulf (C.15) was the most realistic one. In Figure 5.5 it is nice to see how the more recent characters (when realistic) were placed on the right of Beowulf and presented higher comfort. In 2012, the two characters that had the

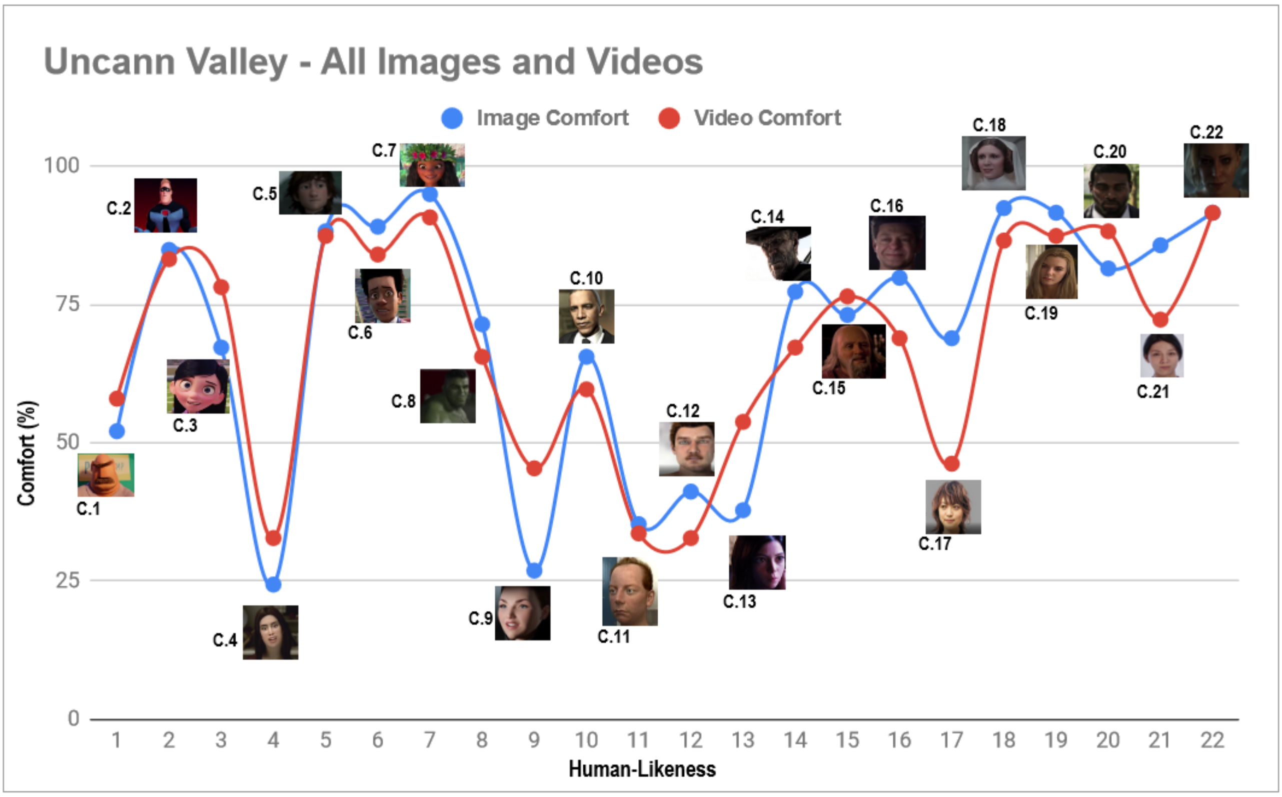


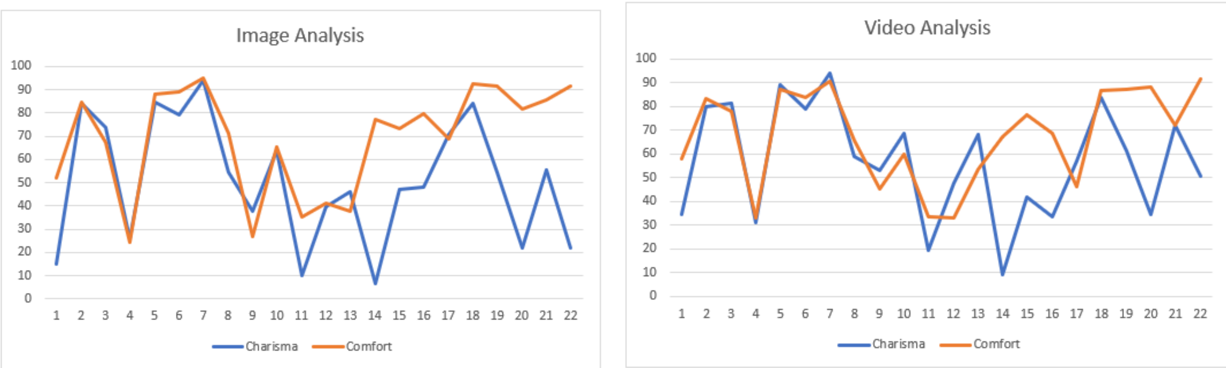
Figure 5.5: All Characters evaluated our approach. Characters 1, 2, 4, 9, 10, 11, 12, 14, 15 and 17 were the 10 evaluated characters in 2012 [26].

highest comfort values were Mr. Incredible (C.2) and Beowulf (C.15) whose values were 88% and 87%, respectively. Even when including the more recent characters, the same two characters (C.2 and C.15) kept high values of comfort; however, the more modern characters still presented higher values in comparison, such as characters 5, 6, 7, 18, 19 and 22.

Correlating Charisma and Comfort

In this section we investigate the correlation between the comfort seen in question Q5 with the charisma given by the answers of question Q4. This was done to assess the influence of characters' charisma on the *Uncanny Valley* effect. For each character, we compute the percentage of people who answered **CHARISMATIC** to characterize the character and correlate these values with comfort. Figure 5.6 contains two charts showing the charisma and comfort obtained for each character in images and videos.

The average value of the characters' charisma in images is 50.87% (standard deviation is 26.45%), while the average in videos is 56.79% (standard deviation is 23.46%). In addition, we use *Pearson's correlation* to measure the relationship between charisma and comfort. The correlation obtained in images is 0.5059 and in videos we obtained a similar value of 0.5029. It is interesting to note that comfort and charisma are directly correlated, as one can expect. With this, we propose the null hypothesis $H0$ defining that comfort and charisma are significantly similar. We performed a *Chi-squared test* with a significance level of 5% to assess the relation between charisma and comfort.



(a) Image Comfort Vs Charisma

(b) Video Comfort Vs Charisma

Figure 5.6: Correlations between charisma and comfort in images (a) and videos (b).

In both analyzes the *p-values* of the comparisons between the mean comfort and charisma were less than 0.05, $2.1E-44$ in the image analysis and $1.8E-26$ in the video analysis, rejecting the null hypothesis. Therefore, **we can confirm that comfort and charisma are directly correlated**.

In addition to charisma, one question is still open: "Is the perceived charisma represented by a character's facial expression?" In addition to the characters' charisma analysis, we measured the facial emotion of virtual characters in order to relate them with their charisma. We used *Openface* [2] [3], a free open source face recognition software that uses *Deep Neural Networks* to capture features and up to 17 *Action Units* [18] (*AU* - facial expressions) on photos and videos. Using *Openface*, we are able to obtain the intensity of each action unit on each image in the interval $[0; 100]$. So, for each virtual character viewed in Figure 5.1, we executed *OpenFace* and processed the *AUs* activated in Happiness as the only positive emotion, and in Fear, Angry and Disgust as the Negative emotions. We average the value of all active *AUs* for each of these mentioned emotions. We included this analysis because we wanted to evaluate if the perceived charisma was affected by a character's facial expression.

Figure 5.7 shows the values of charisma, comfort, happiness and negative emotions for all 22 characters images [4] evaluated in this analysis. As we expected, the facial expressions of almost all characters, as detected by *OpenFace*, seem more neutral than highly negative or positive, not only when analyzing the *OpenFace* result but also when doing a visual inspection. The highest value obtained of happiness was from C.16. Even with that, some characters were classified as charismatic (value close to 90%) and not charismatic (value close to 10%). **It indicates that characters' charisma was not influenced by facial expressions.**

Correlating Familiarity and Comfort

In this section we investigate the correlation between the comfort, asked in question Q5, and the familiarity with the character, given by the answers of question Q3. This was done to assess

[3] I would like to thank VHLab colleague Júlia Melgare for her help in using the *OpenFace* application.

[4] Two characters (C.1 and C.14) were not recognized in *OpenFace* so we did not have information about their *AUs*.

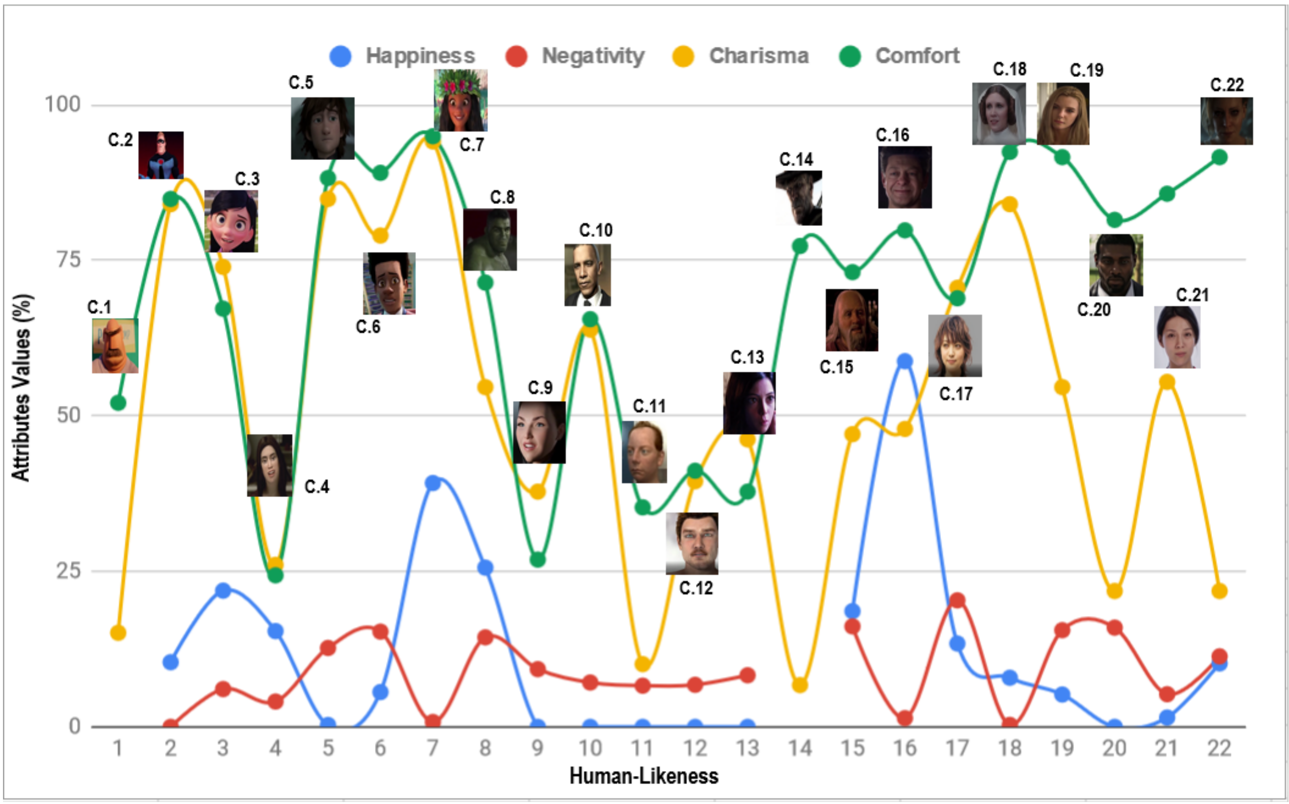
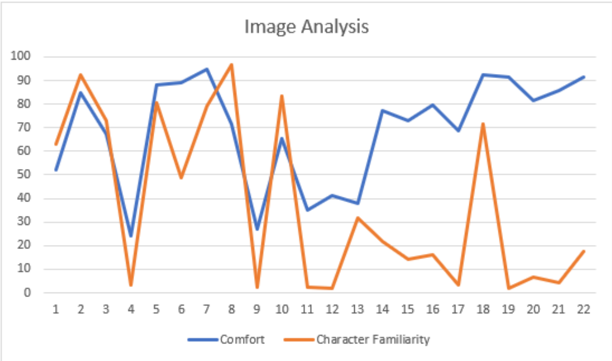


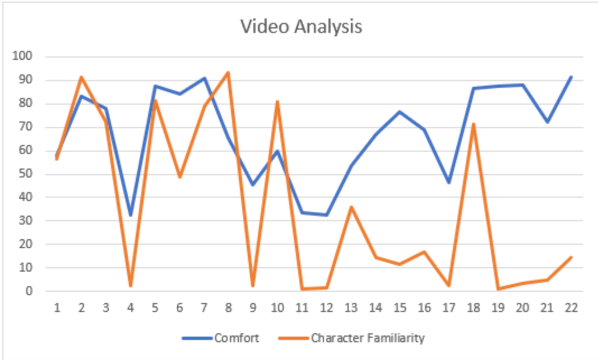
Figure 5.7: The chart shows the Attributes Values obtained in images in (Y-axis) and our order of the characters based on Human Likeness in (X-axis). The Attributes Values used are Charisma (yellow line), Comfort (green line), Happiness (blue line) and Negativity (red line). Each character was represented by a point on each of these lines.

the influence of characters familiarity on the effect of *Uncanny Valley*. As in the charisma section, we calculated the percentage of people who answered "YES" in Q3, but in this case to know if the character was known. Figure 5.8 contains two charts showing the familiarity with the character and the obtained comfort.

The average familiarity with the characters in the images is 37.08%, with a standard deviation of 35.49%, while the average in the videos is 35.79%, with a standard deviation of 35.43%. We use *Pearson's correlation* to measure the relationship between familiarity and comfort. The correlation in the images is 0.35 and in the videos is 0.41, i.e., both variables seem to be weakly correlated. Therefore, in this case, it is also possible to note that the comfort and familiarity with the character are directly correlated, as one would expect, but this correlation is less clear than the correlation between charisma and comfort. We propose the null hypothesis $H0$ defining that comfort and familiarity with the character are similar. As with charisma analysis, a *Chi-square test* with a significance level of 5% was performed to assess the relationship between familiarity and comfort. In both analyzes, the *P-values* of comparisons between comfort and familiarity with character were less than 0.05, $1.5E-123$ in image analysis and $1.4E-123$ in video analysis, rejecting the null hypothesis. Therefore, **we can confirm that comfort and familiarity with character are statistically independent.**



(a) Image Comfort Vs Character Familiarity



(b) Video Comfort Vs Character Familiarity

Figure 5.8: Correlations between comfort and character familiarity in images (a) and videos (b).

5.2.2 Data Obtained in Questionnaire 2

The second questionnaire was answered by 42 participants in our *Research Labs*, where the age varies from 20 to 30 years old, 80% of the participants being male and all undergraduate or graduate students in *Computer Science*. All participants were volunteers in the experiment, and could stop responding if they felt tired or any other problem. People interacted in the environment through a *HP Mixed Reality Headset* [5]. Figure 5.2 illustrates the environment with one of three tested characters in the both evaluated densities. We ask participants to interact with the environment by walking around the room and being as close as possible to the agents. We consider that interactions should last at least 2 minutes, but no one has tried to stop before.

Figure 5.9 indicates the effect of *Uncanny Valley* on the participants' perception when answering the second questionnaire in this analysis after interacting with characters through *VR*. It is possible to see that the non-realistic model (in the center of Figure 5.9) generated low values of comfort in all tested domains. Some factors may explain this event, such as the absence of the eyeball and the cartoon format. In addition, it is interesting to note that the model on the right in Figure 5.9 presented the highest level of comfort. This analysis indicates the *Uncanny Valley* effect worked as expected, based on literature [49], for such three characters. It is important to mention that these three characters were evaluated as all the other characters in this analysis, i.e., only in terms of image, animation, *VR* interaction and organized groups. In this case, we hypothesize that for such tested characters, **we can say that the *VR* interaction does not impact the comfort perception**. The same happens with the characters' density, i.e., results are not impacted by the number of characters in the virtual world. We do not perform statistical analysis on these experiments because of the small amount of interactive characters and participants.

Summarizing Findings Regarding Perception of CG Characters

[5] https://www8.hp.com/us/en/campaigns/mixedrealityheadset/overview.html

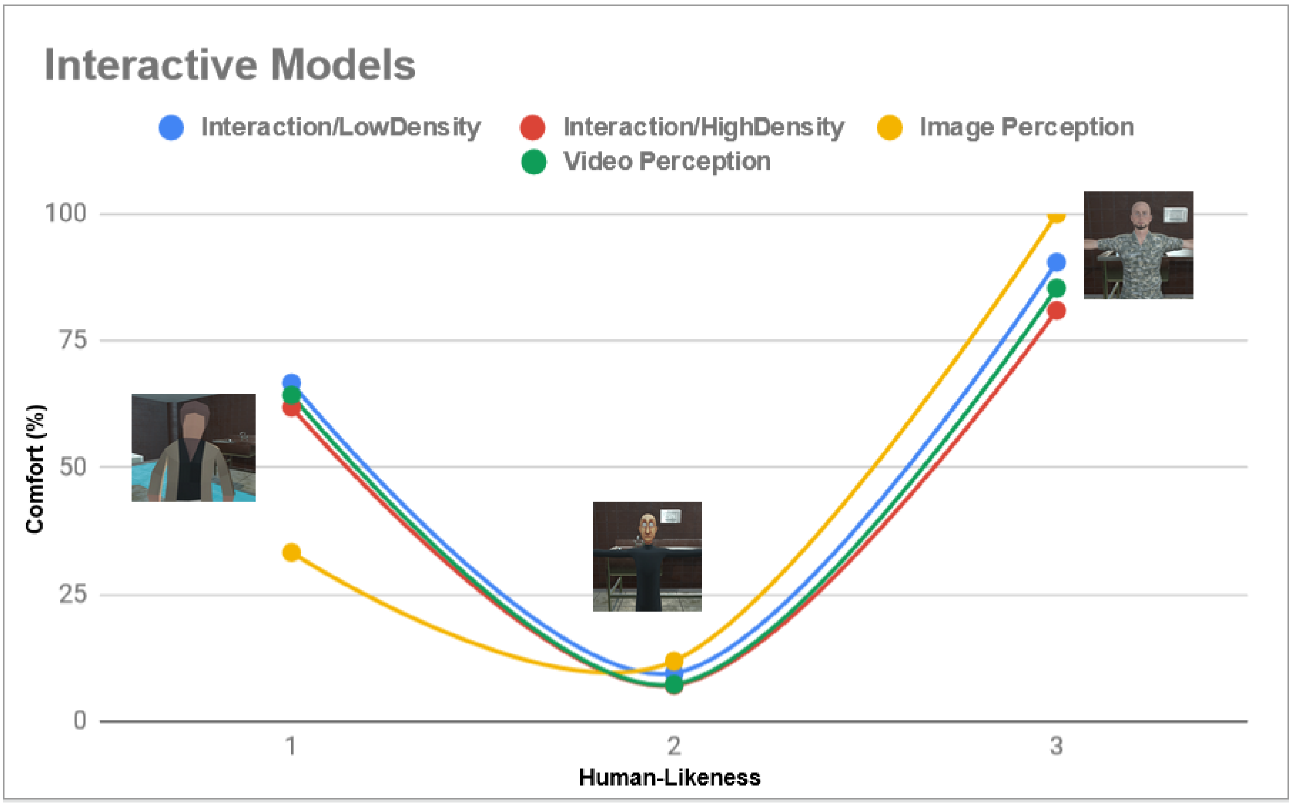


Figure 5.9: Effect of *Uncanny Valley* in the interaction environment.

- Regarding the comparison between Flach's work and our work on the perception of comfort of only the old characters (only the characters used in the work of Flach et al. [26]), results indicated that people in 2012 felt more comforted than 2019 people with respect to these characters. This can be explained by the fact that today we are exposed to graphics with better qualities;

- Regarding the comparison between Flach's work and our work on the perception of comfort of all characters (22 characters in our work, and 12 characters in Flach's work), results indicated that people in 2019 felt more comfortable than the people of 2012 regarding these characters. With that, we can say that nowadays people feel more comfortable with CG characters than in 2012. An indication of this result is the high comfort both in the images and in the videos about the current characters;

- Regarding the comparison between comfort and charisma, results indicated that they are directly correlated.

- Still regarding the charisma, we evaluated whether it was affected by the character's facial expression. Even though most of the characters were evaluated as charismatic, results indicated that the charisma is not affected by the character's facial expression;

- Regarding the comparison between comfort and familiarity with the characters, results indicated that there is a weak correlation between them;

- Regarding the *VR* experiment, results indicated that both the interaction with characters and the density did not impact the perceived comfort.

6. FINAL CONSIDERATIONS

In this work, we performed and evaluate studies on human perception regarding virtual humans and crowds, in order to answer the main research question "What can we learn about human perception of visualizing data obtained in video footage or simulating groups with virtual humans? ". In order to address this aspect, we proceed with three topics of investigations:

i) Perception of People Interactions in Crowds - an interaction model was introduced, based on *Hall's Personal Space*, in which agents/individuals could interact with each other. To analyze the large amount of interaction data (focused on geometric information), we developed visualization of interactions which can be used to find relevant information in both simulations and video output data, like which simulations presented more interactions, or which person, in a given video, was able to interact more, that is, helping with data interpretation and perception. Tests showed that agents were able to interact as intended and following the method discussed in Section 3.1.1. In addition, we considered the proposed visualizations were useful in order to understand simulated data. Using them, we can easily find out which simulations generated more interactions, which agents interacted the most, which personalities generated more interactions, among other aspects. The same happened with video sequences analysis, in which we were able to find the video with more performed interactions and individuals who interacted more or less. Still in the analysis of video sequences, we could see trajectories where people interact as a function of time. For example, in Figure 3.7(b) of Section 3.2.4, we can see the scatter plot of BR-03 video with this interaction behavior. During this video, some individuals walk close to each other, that is, having great possibilities for interactions between them, as they are within the personal space, as proposed by Hall [32]. In addition, our method of generating, finding and visualizing interactions can be useful for game developers, allowing them to generate characters interacting in a more natural way and based on personality models (e.g., using *OCEAN* features). As possible future work on interactions in crowds, some types of perception can be addressed. For example, using synthetic vision to initiate possible interaction before personal space, that is, using communication between individuals, as a type of interaction. Finally, new visualizations would be interesting, capable of providing different information about interactions.

ii) Perceptions of Cultural Features in Crowds - This analysis evaluated people's perceptions of geometric (density, speed, angular variation, and distance) and non-geometric (personality traits and emotions) features through two questions: *i)* "Is human perception about geometric features affected by different viewpoints and avatars?", and *ii)* "Can people perceive cultural features in virtual humans without body and facial expressions?". For this, we proposed and implemented a survey that has been answered by 73 participants through a questionnaire that featured visualizations of scenes taken from videos of the *Cultural Crowds* [20] data set and propose questions regarding variation of visualization parameters. Regarding the results of the first question, in the analysis of the cameras, it was noticed that the cameras point of view influences the human perception (excluding the speed parameters). In particular, the top of view camera had the highest rate among the studied types of camera. In the general analysis of the type of avatars, it was also observed that the choice of one of the types of avatar influences the perception of the parameters. However, neither type (cylinder and humanoid) had a high rate separately. Regarding the second question, w.r.t. cultural properties, it was noted that people can perceive such type pf parameters in virtual humans without body and facial expressions. Interestingly, even without explaining to the participants the concepts of personality or emotion, most of them noted the cultural features expressed by virtual humans, according to our approach. Obviously, this last aspect is much more intangible and can depend on personal aspects, as participants culture and experience. Future work should evolve further research on this aspect.

iii) Perception of CG Characters - In this analysis it was proposed a set of experiments to evaluate how people perceive avatars in the contexts of still images, animations and interactive *VR* scenarios. We tried to answer the following questions: *i)* "Does the exposure to virtual characters, which has been going on for several decades, reduce the Uncanny effect on people's perceptions?", *ii)* "Does the charisma and familiarity with virtual humans correlate to the *Uncanny Valley*?", *iii)* "How does the *Uncanny Valley* effect impact in a Virtual Reality (*VR*) and Interactive Environment?", and finally *iv)* How does the *Uncanny Valley* manifest itself if we have groups of characters and not only one agent?". Regarding the first question, observing Figure 5.5, we can see that the most modern characters presented higher levels of comfort than the characters used in Flach et al.'s work. On the other hand, the average comfort reduces, in the present evaluation, when analysing only the characters evaluated in 2012. In addition to the level of comfort, some older characters had a fall in the order of human likeness, represented in the horizontal axis of each chart. In addition, we measure the difference between the image and video comforts of both Flach's work and all characters evaluated in our approach. From these results, we can see that the average comfort perceived nowadays is higher than the average comfort values of Flach's work, **indicating that people are becoming more comfortable over time with CG characters.** Regarding the second question, in the analysis of charisma, our results **indicate that charisma has a positive correlation with comfort.** However, it was not evident to analyze facial expressions in the context of the characters' charisma, because they purposely have neutral expressions just to focus on geometry and animation rather than emotion. Another set of experiments is required to complete this analysis in a future work. Regarding the familiarity, **we can see that familiarity with the character is also positive related to comfort, even if values are smaller than the correlation with charisma.** Regarding the third and fourth questions, looking at Figure 5.9, we can see that there is an indication of a valley when people perceived and interacted with a virtual human that was not realistic or cartoon. Therefore, comparing with image and video perception, it indicates that **the number of agents or the interactive *VR* environment itself does not impact the final perception of *Uncanny Valley***, but more tests are necessary to have a statistically valid result. As a possible future work, expand the experiments with more participants, with more advanced interactions (conversation and gesture) and focusing mainly on interactive characters that can provide facial animation.

BIBLIOGRAPHY

[1] Alexander, R. M. "A model of bipedal locomotion on compliant legs", *Philosophical Transactions of the Royal Society B*, vol. 338–1284, Oct 1992, pp. 189–198.

[2] Amos, B.; Bartosz, L.; Satyanarayanan, M. "Openface: A general-purpose face recognition library with mobile applications", Technical report, CMU School of Computer Science, 2016, 20p.

[3] Ardeshir, S.; Borji, A. "Egocentric meets top-view", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, May 2018, pp. 1353–1366.

[4] Atkinson, R. C.; Shiffrin, R. M. "Human memory: A proposed system and its control processes". Elsevier, 1968, chap. 2, pp. 89–195.

[5] Atkinson, R. C.; Shiffrin, R. M. "The control of short-term memory", *Scientific American*, vol. 225, Aug 1971, pp. 82–91.

[6] Bernstein, D. "Essentials of Psychology". Cengage Learning, 2018, 864p.

[7] Bicho, A. L. "Da modelagem de plantas à dinâmica de multidões: um modelo de animação comportamental bio-inspirado", Phd thesis, Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, 2009, 114p.

[8] Cai, Z.; Yu, Z. L.; Liu, H.; Zhang, K. "Counting people in crowded scenes by video analyzing". In: Proceedings of the 9th IEEE Conference on Industrial Electronics and Applications, 2014, pp. 1841–1845.

[9] Chandran, A.; Loh, A. P.; Vadakkepat, P. "Identifying social groups in pedestrian crowd videos". In: Proceedings of the 8th International Conference on Advances in Pattern Recognition, 2015, pp. 1–6.

[10] Chattaraj, U.; Seyfried, A.; Chakroborty, P. "Comparison of pedestrian fundamental diagram across cultures", *Advances in Complex Systems*, vol. 12, Mar 2009, pp. 393–405.

[11] Chen, Y. "Agent-based research on crowd interaction in emergency evacuation", *Cluster Computing*, vol. 23–1, Sep 2017, pp. 189–202.

[12] Costa, P. T.; McCrae, R. R. "Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI)". Psychological Assessment Resources, 1992, 101p.

[13] Costa, P. T.; McCrae, R. R. "Neo pi-r: inventario de personalidad neo revisado". TEA Ediciones, 2002, 25p.

[14] Digman, J. M. "Personality structure: Emergence of the five-factor model", *Annual Review of Psychology*, vol. 41, Feb 1990, pp. 417–440.

[15] Dihl, L.; Testa, E. S.; Knob, P.; da Silva, G. L. B.; Favaretto, R. M.; de Alcântara, M. F.; Musse, S. R. "Generating cultural characters based on hofstede dimensions". In: Proceedings of the IEEE Virtual Humans and Crowds for Immersive Environments, 2017, pp. 1–5.

[16] Durupınar, F.; Güdükbay, U.; Aman, A.; Badler, N. I. "Psychological parameters for crowd simulation: From audiences to mobs", *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, Sep 2016, pp. 2145–2159.

[17] Ekman, P.; Friesen, W. V. "Constants across cultures in the face and emotion", *Journal of Personality and Social Psychology*, vol. 17, Jan 1971, pp. 124–129.

[18] Ekman, P.; Friesen, W. V. "Measuring facial movement", *Environmental Psychology and Nonverbal Behavior*, vol. 1, Sep 1976, pp. 56–75.

[19] Ennis, C.; Peters, C.; O'Sullivan, C. "Perceptual effects of scene context and viewpoint for virtual pedestrian crowds", *ACM Transactions on Applied Perception*, vol. 8, Feb 2011, pp. 1–22.

[20] Favaretto, R. M.; Dihl, L.; Barreto, R.; Musse, S. R. "Using group behaviors to detect hofstede cultural dimensions". In: Proceedings of the IEEE International Conference on Image Processing, 2016, pp. 2936–2940.

[21] Favaretto, R. M.; Dihl, L.; Musse, S. R.; Vilanova, F.; Costa, A. B. "Using big five personality model to detect cultural aspects in crowds". In: Proceedings of the 30th Brazilian Symposium Conference on Graphics, Patterns and Images, 2017, pp. 223–229.

[22] Favaretto, R. M.; Dihl, L. L.; Musse, S. R. "Detecting crowd features in video sequences". In: Proceedings of the 29th Brazilian Symposium Conference on Graphics, Patterns and Images, 2016, pp. 201–208.

[23] Favaretto, R. M.; dos Santos, R. R.; Musse, S. R.; Vilanova, F.; Costa, A. B. "Investigating cultural aspects in the fundamental diagram using convolutional neural networks and virtual agent simulation", *Computer Animation and Virtual Worlds*, vol. 30, May 2019, pp. 1–12.

[24] Favaretto, R. M.; Knob, P.; Musse, S. R.; Vilanova, F.; Costa, Â. B. "Detecting personality and emotion traits in crowds from video sequences", *Machine Vision and Applications*, vol. 30, Oct 2018, pp. 1–14.

[25] Favaretto, R. M.; Musse, S. R.; Costa, A. B. "Emotion, Personality, and Cultural Aspects in Crowds". Springer, 2019, 178p.

[26] Flach, L. M.; de Moura, R. H.; Musse, S. R.; Dill, V.; Pinho, M. S.; Lykawka, C. "Evaluation of the uncanny valley in cg characters". In: Proceedings of the Brazilian Symposium on Computer Games and Digital Entertainment, 2012, pp. 108–116.

[27] Goldberg, L. R. “An alternative description of personality: the big-five factor structure”, *Journal of Personality and Social Psychology*, vol. 59, Dec 1990, pp. 1216–1229.

[28] Gorbova, J.; Lüsi, I.; Litvin, A.; Anbarjafari, G. “Automated screening of job candidate based on multimodal video processing”. In: Proceedings of the Conference on Computer Vision and Pattern Recognition, 2017, pp. 29–35.

[29] Gou, L. “Visualizing personality traits derived from social media”. In: Proceedings of the IEEE Visualization, Workshop Personal Visualization: Exploring Data in Everyday Life, 2015, pp. 5.

[30] Gregory, R. L. “Perceptions as hypotheses”, *Philosophical Transactions of the Royal Society B*, vol. 290, Jul 1980, pp. 181–197.

[31] Hall, C. S.; Lindzey, G.; Campbell, J. B. “Theories Of Personality”. John Wiley & Sons, 1998, 768p.

[32] Hall, E. T. “The Hidden Dimension”. New York Anchor, 1966, 215p.

[33] Hodgins, J.; Jörg, S.; O’Sullivan, C.; Park, S. I.; Mahler, M. “The saliency of anomalies in animated human characters”, *ACM Transactions on Applied Perception*, vol. 7, Jul 2010, pp. 22.

[34] Hofstede, G. “Culture’s consequences: Comparing values, behaviors, institutions and organizations across nations”. Sage Publications, 2001, 616p.

[35] Hyde, J.; Carter, E. J.; Kiesler, S.; Hodgins, J. K. “Evaluating animated characters: Facial motion magnitude influences personality perceptions”, *ACM Transactions on Applied Perception*, vol. 13, Feb 2016, pp. 8.

[36] Jacques Junior, J.; Musse, S. R.; Jung, C. “Crowd analysis using computer vision techniques”, *IEEE Signal Processing Magazine*, vol. 27, Sep 2010, pp. 66–77.

[37] Jimenez, J.; Zsolnai, K.; Jarabo, A.; Freude, C.; Auzinger, T.; Wu, X.-C.; von der Pahlen, J.; Wimmer, M.; Gutierrez, D. “Separable subsurface scattering”. In: Proceedings of the Computer Graphics Forum, 2015, pp. 188–197.

[38] John, O. P. “The Big Five factor taxonomy: Dimensions of personality in the natural language and in questionnaires”. New York, NY: 66–100, 1990, chap. 4, pp. 66–100.

[39] Kätsyri, J.; Förger, K.; Mäkäräinen, M.; Takala, T. “A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness”, *Frontiers in Psychology*, vol. 6, Apr 2015, pp. 1–16.

[40] Kätsyri, J.; Mäkäräinen, M.; Takala, T. “Testing the ‘uncanny valley’ hypothesis in semirealistic computer-animated film characters: An empirical evaluation of natural film stimuli”, *International Journal of Human-Computer Studies*, vol. 97, Jan 2017, pp. 149–161.

[41] Kok, V. J.; Lim, M. K.; Chan, C. S. "Crowd behavior analysis: A review where physics meets biology", *Neurocomputing*, vol. 177, Feb 2016, pp. 342–362.

[42] Kullu, K.; Güdükbay, U.; Manocha, D. "Acmics: an agent communication model for interacting crowd simulation", *Autonomous Agents and Multi-Agent Systems*, vol. 31, May 2017, pp. 1403–1423.

[43] Lala, D.; Thovuttikul, S.; Nishida, T. "Towards a virtual environment for capturing behavior in cultural crowds". In: Proceedings of the 6th International Conference on Digital Information Management, 2011, pp. 310–315.

[44] Lamer, S. A.; Sweeny, T. D.; Dyer, M. L.; Weisbuch, M. "Rapid visual perception of interracial crowds: Racial category learning from emotional segregation", *Journal of Experimental Psychology: General*, vol. 147, May 2018, pp. 683–701.

[45] Lord, W. "Neo Pi-R – A Guide to Interpretation and Feedback in a Work Context". Hogrefe Ltd, 2007, 30p.

[46] Ma, L.; Deng, Z. "Real-time facial expression transformation for monocular rgb video". In: Proceedings of the Computer Graphics Forum, 2018, pp. 470–481.

[47] MacDorman, K. F.; Chattopadhyay, D. "Reducing consistency in human realism increases the uncanny valley effect; increasing category uncertainty does not", *Cognition*, vol. 146, Jan 2016, pp. 190–205.

[48] Marčetić, D.; Maleš, L.; Ribarić, S. "Crowd motion pattern detection at the microscopic level". In: Proceedings of the 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics, 2019, pp. 1093–1098.

[49] Mori, M. "Bukimi no tani [the uncanny valley]", *Energy*, vol. 7, Jan 1970, pp. 33–35.

[50] Mustafa, M.; Guthe, S.; Tauscher, J.-P.; Goesele, M.; Magnor, M. "How human am i?: Eeg-based evaluation of virtual characters". In: Proceedings of the Conference on Human Factors in Computing Systems, 2017, pp. 5098–5108.

[51] Mustafa, M.; Magnor, M. "Eeg based analysis of the perception of computer-generated faces". In: Proceedings of the 13th European Conference on Visual Media Production, 2016, pp. 4.

[52] Plaisant, C.; Milash, B.; Rose, A.; Widoff, S.; Shneiderman, B. "Lifelines: visualizing personal histories". In: Proceedings of the Conference on Human Factors in Computing Systems, 1996, pp. 221–227.

[53] Runions, A.; Fuhrer, M.; Lane, B.; Federl, P.; Rolland-Lagan, A.-G.; Prusinkiewicz, P. "Modeling and visualization of leaf venation patterns", *ACM Transactions on Graphics*, vol. 24, Jul 2005, pp. 702–711.

[54] Saifi, L.; Boubetra, A.; Nouioua, F. "An approach for emotions and behavior modeling in a crowd in the presence of rare events", *Adaptive Behavior*, vol. 24, Nov 2016, pp. 428–445.

[55] Schacter, D.; Gilbert, D.; Wegner, D.; Hood, B. M. "Psychology: European Edition". Macmillan International Higher Education, 2011, 808p.

[56] Schwind, V.; Wolf, K.; Henze, N. "Avoiding the uncanny valley in virtual character design", *Interactions*, vol. 25, Aug 2018, pp. 45–49.

[57] Seymour, M.; Riemer, K.; Kay, J. "Interactive realistic digital avatars-revisiting the uncanny valley". In: Proceedings of the 50th Hawaii International Conference on System Sciences, 2017, pp. 547–556.

[58] Solmaz, B.; Moore, B. E.; Shah, M. "Identifying behaviors in crowd scenes using stability analysis for dynamical systems", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, Oct 2012, pp. 2064–2070.

[59] Yang, F.; Shabo, J.; Qureshi, A.; Peters, C. "Do you see groups?: The impact of crowd density and viewpoint on the perception of groups". In: Proceedings of the 18th International Conference on Intelligent Virtual Agents, 2018, pp. 313–318.

[60] Yantis, S. "Stimulus-driven attentional capture and attentional control settings", *Journal of Experimental Psychology: Human Perception and Performance*, vol. 19(3), Jun 1993, pp. 676––681.

[61] Zell, E.; Aliaga, C.; Jarabo, A.; Zibrek, K.; Gutierrez, D.; McDonnell, R.; Botsch, M. "To stylize or not to stylize?: the effect of shape and material stylization on the perception of computer-generated faces", *ACM Transactions on Graphics*, vol. 34, Oct 2015, pp. 184.

[62] Zeng, W.; Fu, C.-W.; Arisona, S. M.; Schubiger, S.; Burkhard, R.; Ma, K.-L. "A visual analytics design for studying crowd movement rhythms from public transportation data". In: Proceedings of the Conference and Exhibition on Computer Graphics and Interactive Techniques in Asia, 2016, pp. 4.

[63] Zhao, J.; Gou, L.; Wang, F.; Zhou, M. "Pearl: An interactive visual analytic tool for understanding personal emotion style derived from social media". In: Proceedings of the IEEE Conference on Visual Analytics Science and Technology, 2014, pp. 203–212.

[64] Zibrek, K.; Kokkinara, E.; McDonnell, R. "The effect of realistic appearance of virtual characters in immersive environments-does the character's personality play a role?", *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, Apr 2018, pp. 1681–1690.

APPENDIX A – PAPERS WRITTEN DURING THE MASTER'S DEGREE

- KNOB, P.; ARAUJO, V. F. A.; FAVARETTO, R. M.; MUSSE, S. R. "Visualization of Interactions in Crowd Simulation and Video Sequences." Proceedings of the 17th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames). IEEE, 2018. Qualis B2 *(Published Paper)*

- ARAUJO, V. F. A.; FAVARETTO, R. M.; KNOB, P.; MUSSE, S. R.; VILANOVA, F.; COSTA, A. B. "How much do you perceive this?: An analysis on perceptions of geometric features, personalities and emotions in virtual humans." Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents. ACM, 2019. Qualis B1 *(Published Paper)*

- DIAS, W.; FELIX, L.; MELGARÉ, J.; ARAUJO, V. F. A.; KNOB, P.; MUSSE, S. R. "How do true eyes really move in video sequences?" Proceedings of the 18th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames). Workshop G2 Undergraduates. 2019. *(Published Paper)*

- FAVARETTO, R. M.; COSTA, A. B.; MUSSE, S. R. "Emotion, personality and cultural aspects in crowds: towards a geometrical mind". 2019. **(Chapter 7 Contributor: Detecting Personality and Emotion Traits)** *(Published Chapter)*

- FAVARETTO, R. M.; COSTA, A. B.; MUSSE, S. R. "Emotion, personality and cultural aspects in crowds: towards a geometrical mind". 2019. **(Chapter 10 Contributor: Video Analysis Dataset And Applications)** *(Published Chapter)*

- ARAUJO, V. F. A.; FAVARETTO, R. M.; KNOB, P.; MUSSE, S. R.; VILANOVA, F.; COSTA, A. B. "How much do we perceive geometric features, personalities and emotions in virtual humans?". SN Computer Science. 2020. *(Submitted Paper)*

- ARAUJO, V. F. A.; MELGARÉ, J.; GEISS, R.; PINHO, M.; MUSSE, S. R. "How do we perceive Avatars? A Perception Analysis in Still Images, Animations and VR Scenarios". IEEE Transactions on Visualization and Computer Graphics (TVCG). 2020. *(Submitted Paper)*

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