



Cultural behaviors analysis in video sequences

Rodolfo Migon Favaretto¹ · Victor Flavio de Andrade Araujo¹ · Felipe Vilanova² · Angelo Brandelli Costa² · Soraia Raupp Musse¹

Received: 22 September 2020 / Revised: 18 March 2021 / Accepted: 9 June 2021 / Published online: 4 July 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

In this paper, we investigate the cultural aspects of different populations from video sequences. For that, we proposed a model that considers a series of characteristics of the pedestrians and the crowd, such as distances and speeds and performs the mapping of these characteristics in personalities, emotions, and cultural aspects. The model called *Big4GD* consists of four dimensions of geometric characteristics and seeks to describe the behavior of pedestrians and groups in the crowd. We performed a study of group behavior in a controlled experiment and focused on differences in two attributes that vary across cultures: (walking speed and personal distance) in three countries (India, Brazil, and Germany). We use the Fundamental Diagram theory that determines the relationship between the density and speed of individuals. We use Computer Vision methods to detect and track individuals through video sequences by generating their positions and speeds as a function of time. With these data, we analyze emergent walking speeds and densities while considering the personal distance of each individual and the neighbor in front of him/her. Our results show that human behavior is more similar in highly dense populations, i.e., individuals behave like a mass when presented with limited free personal space. The opposite result is also relevant: cultural differences can be observed at low and moderate densities, and such assumptions can be applied to computational interfaces and simulations, games, and movies. Besides, we present *GeoMind*, a software we developed to detect a series of characteristics from pedestrians. We also performed a practical case-study using *GeoMind* focusing on event detection in video sequences.

Keywords Crowd analysis · Cultural aspects · Computer Vision · Fundamental Diagram

1 Introduction

According to Cambridge Dictionary [42], culture is defined as “the way of life, especially the general customs and beliefs, of a particular group of individuals at a particular time.” Cultural aspects are widely studied in many research fields. In the work performed by [47], the authors compared relationship-specific social touching patterns between one East Asian and one European culture. In [41], authors explored the perception of emotion appraisal dimensions based on speech prosody in a cross-cultural setting, considering Australia and

India. In this paper, we are focused on video understanding and specifically concerning the context of cultural aspects observed in groups of pedestrians.

Hofstede [30,31] defined his cultural dimensions based on questionnaires completed by individuals from several countries. In a previous work [19], we proposed a mapping from geometrical data (positions and orientations) to Hofstede’s cultural dimensions, using video sequences. Besides, in [20,22], we compared Hofstede’s cultural measures with personality traits [17] using the Big Five personality model: 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”); and Neuroticism (“how prone an individual is to psychological distress”) [35].

✉ Rodolfo Migon Favaretto
rodolfo.favaretto@edu.pucrs.br

¹ Virtual Humans Simulation Laboratory (VHLab) - Graduate Programme in Computer Science, Pontifical Catholic University of Rio Grande do Sul (PUCRS), Av Ipiranga, 6681, Porto Alegre, RS, Brazil

² Graduate Programme in Psychology (PUCRS), Av Ipiranga, 6681, Porto Alegre, RS, Brazil

In previous papers, we worked with videos collected from different countries to study Hofstede's cultural dimensions [30] and the Big Five personality traits [17] in the context of cultural differences observed among populations. A major challenge discussed previously [21] concerns the investigation of culture in videos of different situations. For example, it would not be appropriate to compare a video with some carefree people strolling in the mall with a video where there are many people at a train station, hasty, and preoccupied with the time, since the context is different. It is expected that such individuals would behave differently, as the environment and contexts filmed are completely different. Thus, from such situations, it is difficult to establish whether the differences we found are related to cultural aspects, once differences in individual behavior can be explained by other variables (e.g., the context). As a contribution to this line of research, this paper describes a cultural analysis of videos of groups of individuals from different countries while studying these populations in the same context.

Our main hypothesis is that personal space and walking speeds observed in pedestrian videos of varied densities can be used as an indicator of the presence of cultural behaviors. As stated above, we investigated cultural aspects from a same-task context to exclude other variables, such as the physical environment. We use the Fundamental Diagram (FD) proposed by [14], in which individuals from three analyzed countries are required to execute the same task. The FDs originally proposed in use in traffic planning guidelines [13] are diagrams used to describe the relationship between three parameters: i) the density of individuals (the number of individuals per sqm), ii) speed (in meters/second), and iii) flow (time evolution) [54]. Seyfried et al. [44] discussed the empirical relationship between the density and velocity of pedestrian movement in a unidirectional flow. Other works [9,10,23] have addressed bi-directional and multi-directional flows observed at corridors and crossings.

In the work proposed by Zhang [55], the authors adapted FDs to describe the relationship between pedestrian flow and density, which is associated with various phenomena of self-organization in crowds (e.g., pedestrian lanes and jams) such that when the density of individuals is very high, the crowd stops moving. This is not the first work that used FDs related to cultural aspects. Chattaraj and his collaborators [14] suggest that cultural and population differences can also change the speed, density, and flow of individuals. In their work [14], the influence of culture on the trajectory of pedestrians is investigated from fundamental diagrams.

In this paper, we investigate cultural aspects (differently from [14]) observed from FD videos using such information as personal space (explained in the next section) and individual speeds. We used data collected from India and Germany as presented by [14] and conducted an experiment in Brazil based on the same rules and constraints. Next, we

compared the three populations while performing the same task to determine their cultural differences. The organization of this paper is defined as follows: Section 2 presents an overview of the *Big-Four Geometrical Dimensions* model. In Sect. 3, we discuss some related work while in Sect. 4 we described the proposed methodology. Section 5 presents the software *GeoMind*, Section 6 presents the evaluations we performed, and Section 7 draws some final discussion and presents future work.

2 Overview

The motivation for the research in this area starts from the need to have computational tools and models that are capable of extracting characteristics from real pedestrians and crowds, benefiting several other areas of knowledge, e.g., the computational simulation. The relevance of this research is justified due to its diverse applications, such as physical space planning, entertainment, and security management, among others. These areas seek to consider regional and cultural aspects, but usually, this is done with empirical knowledge, there is no computational method or model to help in this task. In the area of physical space management and planning of environments, regional and cultural behaviors and habits should be considered [28,40]. For example, in Arabian cities, the provision of public and private spaces is, for the most part, very different from the European cities, according to a study performed by [28].

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

To address this problem, we proposed the development of a computational model that allows the extraction of pedestrian and crowd characteristics based on their behaviors, manifestation in space/time, being able to identify cultural, personality, and emotional aspects and behaviors that differ from one country to another. Figure 1 shows an overview of the features mapping of the proposed approach. Directional arrows indicate which features are used as inputs to calcu-

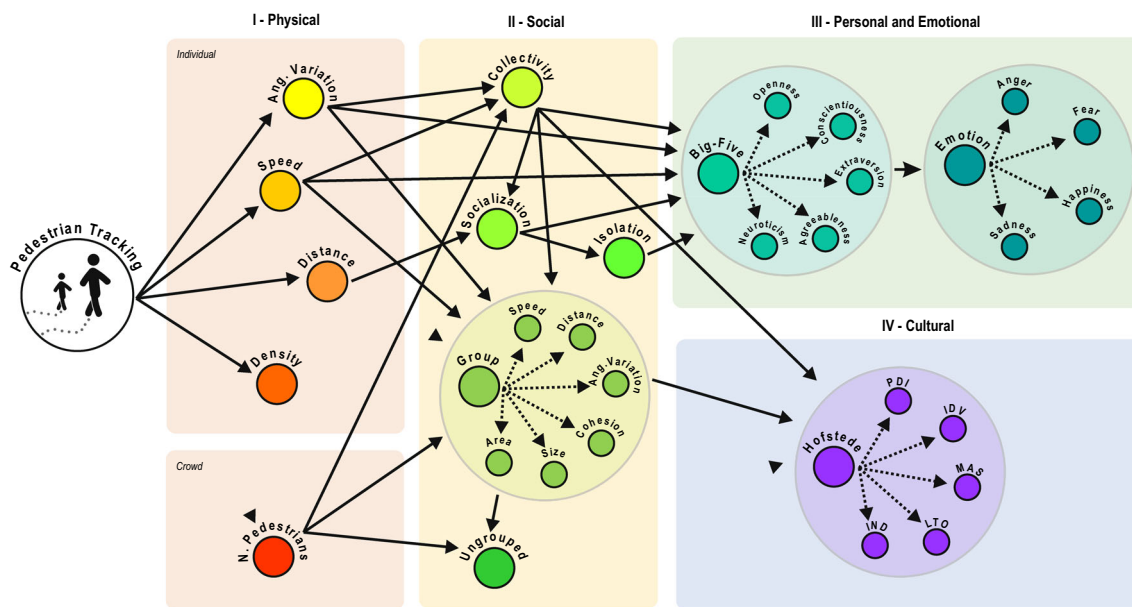


Fig. 1 Our Big-Four Geometrical Dimensions: features mapping from the proposed approach divided into four dimensions: *I-Physical*, *II-Social*, *III-Personal and Emotional*, and finally, *IV-Cultural*. Directional arrows indicate which features are used as inputs to calculate others

late others, for example, *distance* is used as input to calculate *socialization* and to detect *groups*.

We proposed four different dimensions of features to characterize pedestrians organized in groups/crowds in video sequences. These dimensions, named as **Big-Four Geometrical Dimensions** or just **Big4GD** model, illustrated in Fig. 1, are:

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

In addition to the features mapping, we present in Fig. 2 a high-level flowchart describing our approach. It shows how our method is organized and, together with Figure 1, helps us to better understand how it works. Each part of the flowchart is described later in the paper.

Also, each of the dimensions presented in Fig. 1 is detailed later, in Sect. 4. Next we present some related work regarding speed and distance correlated to cultural behaviors.

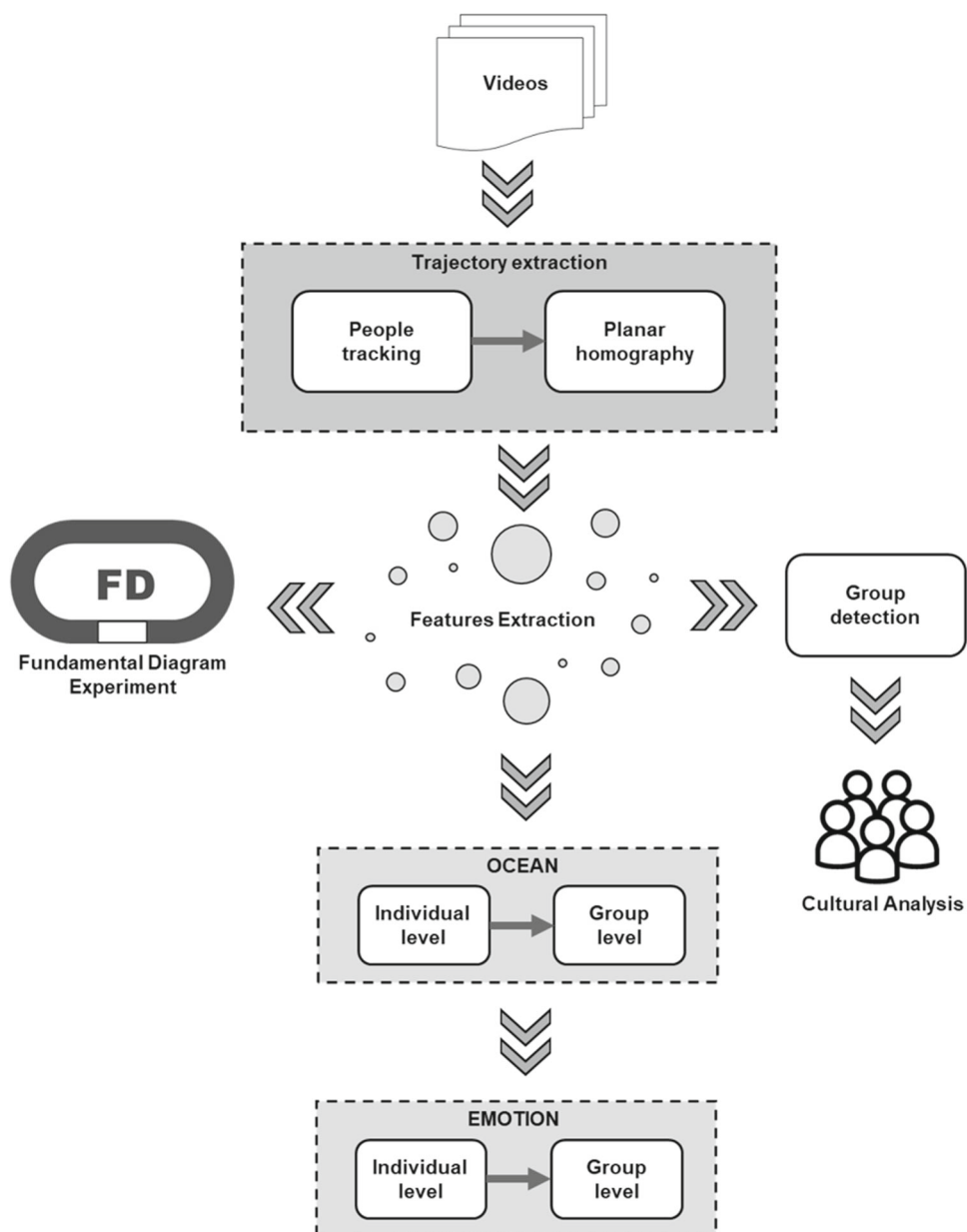
3 Related work

This section addresses some background about individual, groups, and crowd behaviors detection techniques (Section 3.1) and related work to correlate walking speed and personal distance analyses with cultural behaviors (Section 3.2).

3.1 Humans, groups, and crowd behaviors detection techniques

The pace of life is defined as the rate, speed, and “relative rapidity or density of experiences, meanings, perceptions, and activities” [53]. Pace of life has been measured as individuals’ mean walking speeds collected from streets, as the mean length of time postal workers take to complete a standard request for stamps and as the accuracy of street clocks [34]. These components vary cross-culturally and are related to several community-related characteristics, such as average annual high temperatures, gross domestic product per capita, and the collectivism index of a given country [16,34]. The importance of these components has been previously stressed, as they affect not only subjective well-being but also a country’s economic conditions such that faster-paced places tend to be more economically productive [29]. The method used to measure the pace of life components can be enhanced through the use of modern techniques. Walking speed, for instance, was originally measured by observing and timing male and female walking speeds over a distance of 60 feet [34]. Through the use of contemporary video-analytic meth-

Fig. 2 Flowchart of our approach: a high-level view from the proposed method



ods, it is possible to enhance the analysis of walking speeds in distinct settings while avoiding imprecision resulting from the human factor. In this sense, by improving the measurement of walking speeds, we can more feasibly measure the cultural characteristics of crowds.

Proxemics is the study of how an individual perceives and uses his or her interpersonal space [25]. Proxemics focuses on the organization of spaces in houses and buildings (e.g., the arrangement of furniture or the use of doors) and on the design of towns, but it focuses primarily on out-of-awareness distance settings [26]. An individual maintains a specific amount of personal space to regulate social intimacy and to control sensory exposure. When positioned at closer distances to others, there is a higher probability of experiencing

increased levels of visual, tactile, auditory, and olfactory stimulation [27], which can cause individuals to feel intruded upon and which can elicit psychophysiological responses to an event [18]. The area surrounding a person's body into which intruders may not enter is called personal space [45].

Personal space refers to the distance that an individual prefers to distance himself or herself from the others within a given setting. Personal space serves two main functions: it communicates the formality of the relationship between interactants and protects an individual from experiencing potentially psychologically and physically uncomfortable social encounters by regulating the quantity and quality of sensory stimulation experienced [1]. According to Baxter, individuals of various cultural backgrounds differ concern-

ing their personal space needs [4]. These differences reflect cultural norms that shape perceptions of space and guide the use of distance in different societies [27]. Different perceptions often lead to different definitions of what constitutes an inappropriate interaction distance, which may lead to miscommunication when individuals from different cultures attempt to interpret one another's spatial behaviors [1].

Originally, Hall [27] distinguished between contact and noncontact cultures. Contact cultures tend to exhibit closer interpersonal distances and to engage in more touching while members of noncontact cultures tend to maintain greater interpersonal distances and to engage in touching less often. According to Hall [27], it is possible to group these characteristics geographically, with Arabian, Latin American, and southern European countries being contact cultures and with Asian, North American, and Northern European being non-contact cultures. Interpersonal distance has also been categorized [27] about public distance maintained daily between unknown individuals on a street; social distance maintained during formal interactions; personal distance maintained during interactions between friends; and intimate distance maintained in close relationships [46].

3.2 Previous approaches in cultural behaviors

This section describes some approaches regarding cultural behaviors analysis. A great study related to personal distance was conducted in 42 countries [46]. Participants were asked to complete a visual survey on the amount of distance that they would need to maintain to feel comfortable when interacting with: a) a stranger, b) an acquaintance, and c) a close relation. The authors, in turn, evaluated projected metric distances for a) social, b) personal, and c) intimate distance. Users answered on a 0 – 220cm scale anchored by two human-like figures. The authors reported significant variability in the personal distance across countries from the three different interaction conditions.

From the study of several countries, Sorokowska and colleagues [46] drew different conclusions for different cultures. As the use of objective measures of personal distance has been encouraged in the literature [1], methods of interactive analysis may be the most appropriate not only for further developing the new potential categorization of cultures but also to design objects or to implement changes in the real world. Public transportation, for example, can be improved through an analysis of personal distance in different countries, as the invasion of personal distance on trains elicits psychophysiological responses of stress [18]. Another application concerns human robots [52]. As robots mustn't invade the personal space of their users, the configuration of distances may benefit from studies of daily preferred interpersonal distances for different countries.

Considering group behaviors, some work seeks to detect and track human groups in crowds [36,39,43]. In [43], the approach is motivated by the collective behavior of individuals motivated by sociological studies. The authors design a tracking framework that is capable of handling short-term occlusions and producing stable trajectories. Human groups are discovered by clustering trajectories of individuals in an agglomerative fashion. In the approach from [39], the authors combine motion and spatial information with low-level descriptors to be robust to situations such as partial occlusions. In [36], the authors proposed a method to detect and track interacting people by employing a framework based on a Social Force Model (SFM).

Some recent work presents approaches in crowd level [11,12,37]. In [11], the authors propose a light-weight and fast fully-convolutional neural network to learn a regression model for crowd counting in images acquired from drones. [12] introduces a crowd detection method for drone safe landing. In the work presented in [37], the authors proposed an approach to detect cultural aspects from crowds using Convolutional Neural Networks (CNN).

In this paper, we investigate the cultural aspects of individuals in analyzing FDs of personal distance and walking speeds for three different countries: India, Brazil, and Germany. We also compared personal distance results obtained through our approach with the results of Sorokowska et al. [46] experiment.

4 Big-four geometrical dimensions

In this section, we present the four dimensions (illustrated in Fig. 1) of the features we proposed to characterize pedestrians from video sequences. Section 4.1 describes the first (*I – Physical*) and second (*II – Social*) dimensions, while Section 4.2 describes the third (*III – Personal and Emotional*) and Section 4.3 the fourth (*IV – Cultural*) dimension.

4.1 First and second dimensions: tracking and data extraction

In this section, we describe the initial steps of the *Big-Four Geometrical Dimensions* model: tracking process and data extraction. The initial detection of people in the video is performed through the real-time object detector proposed by [50]. The tracking input parameters are the initial positions of people's heads. After the tracking step, a set of trajectories, in image coordinates is obtained. This step can be performed by any tracker since the tracking process is not the main contribution of this work (we used the method proposed by [5]). The next step is to correct the perspective of the image and obtain the parameters in world coordinates, for this a homographic planar projection is performed.

Once the trajectories in world coordinates are obtained, the following information is computed for each pedestrian i , at each timestep: *i*) 2D position x_i (meters); *ii*) speed s_i (meters/frame); *iii*) angular variation α_i (degrees) w.r.t. a reference vector $\mathbf{r} = (1, 0)$; *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 individuals, we computed:

$$\phi_i = \sum_{j=0}^{n-1} \gamma e^{(-\beta \varpi(i,j)^2)}, \quad (1)$$

and the collectivity between two individuals is 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 people i and j , and w_1 and w_2 are constants that should regulate the offset in meters and radians. We have used $w_1 = 1$ and $w_2 = 1$. Thus, values for $\varpi(i, j)$ are included in interval $0 \leq \varpi(i, j) \leq 4.34$. $\gamma = 1$ is the maximum collectivity value when $\varpi(i, j) = 0$, and $\beta = 0.3$ is empirically defined as a decay constant. Hence, ϕ_i is a value in the interval $[0; 1]$. \oplus

To compute the socialization level ϑ , we 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 Euclidean distance from a person i to others $\bar{d}_{i,j}$ and the number of people in the Social Space¹ according to Hall's proxemics [27] around the person n_i). In addition, the network has 10 hidden layers and 2 outputs (the probabilities of socialization and non-socialization). The final accuracy was 96%. We used 16.000 samples (70% of training and 30% of validating). Once we get the socialization level ϑ_i , we compute the isolation level $\varphi_i = 1 - \vartheta_i$, that corresponds to its inverse. For more details about how these features are obtained, please refer to [20,21]. For each individual i in a video, we computed the average for all frames and generate a vector \mathbf{V}_i of extracted data where $\mathbf{V}_i = [x_i, s_i, \alpha_i, \varphi_i, \vartheta_i, \phi_i]$. In the next section, we describe how these features are mapped into personality dimensions.

Regarding groups, to define that two pedestrians i and j belong to the same group, they have to attend to three conditions:

- **Condition 1:** if $d(\vec{X}_i, \vec{X}_j) \leq 1.2$ meters. This distance was defined based on the interpersonal space of the proxemics distances proposed by [27];
- **Condition 2:** if $\alpha(\alpha_i, \alpha_j) \leq 15^\circ$. It was empirically defined that the difference in orientation between the two pedestrians could not be greater than 15° ; and

- **Condition 3:** if $s(s_i, s_j) \leq \beta \max\{s_i, s_j\}$. The speed difference between the two pedestrians can not be greater than β , where $\beta = 5\%$ empirically defined.

At first, as soon as a group is detected, it is identified as a group *temporary*, that is, a group that is still unstable. This group becomes a *permanent* group if it maintains its formation (without the input or output of people) for at least 10% of the number of frames of the video. After this step, a series of information about the permanent groups are computed, such as the number G^f of groups, mean speed \bar{s}_g (in meters/second) among all the members of g , mean angular variation $\bar{\alpha}_g$ (in degrees) among all the members of g , group area A_g and group cohesion C_g . All this information is described in [21].

In this paper, we perform an experiment regarding the Fundamental Diagram, which is presented in detail in Sect. 6. In the experiment, we focused on some of the features from the *I - Physical* dimension: *i*) speed, *ii*) distance, and *iii*) density. As an example from the *II - Social* dimension, Figure 3 shows a frame of a video of China with some groups detected using our approach: the group highlighted in blue is a temporary group. This group was detected some frames ago when that three pedestrians entered in the video. If those pedestrians remain in the formation for at least 10% of the frames of the video, they will become a permanent group (as the other five groups highlighted in red in Fig. 3).

4.2 Third dimension: detection of personality and emotion traits

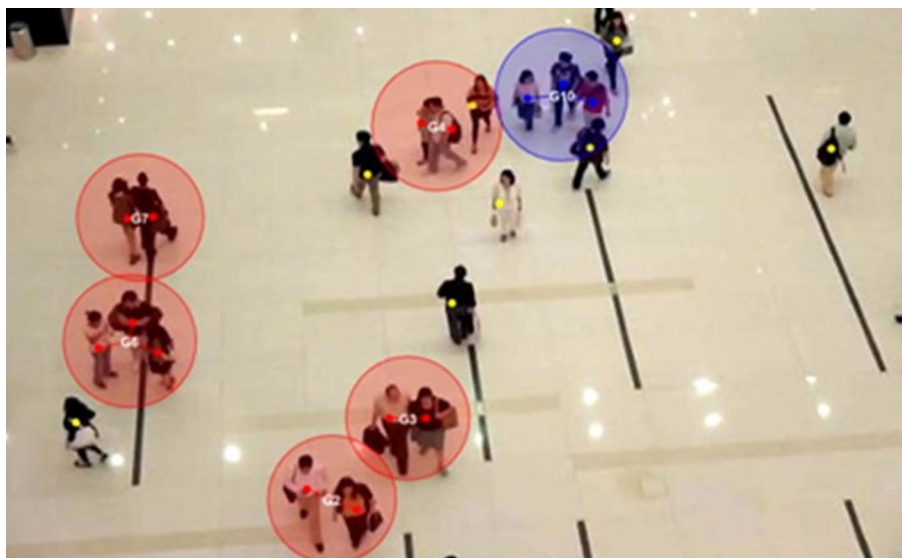
This section presents the proposed methodology to detect personality and basic emotion characteristics of crowds in video sequences, features from the *III - Personal and Emotional* dimension. Firstly, individuals are detected and tracked, then such information is mapped to OCEAN dimensions, used to find out personality and emotion in videos, based on OCC emotion model. Our model presents three main steps as following: *i*) video data extraction, *ii*) personality, and *iii*) emotion analysis.

The first step aims to obtain the individual trajectories from observed pedestrians in real videos. Using these trajectories, we detect groups and extract data that are useful for the second step, which is responsible for the personality mapping of pedestrians, as described in Sect. 4.1 (using the vector \mathbf{V}_i). Once we have concluded the second step, we have enough information to follow with the third step, which consists of emotion detection of individuals and groups according to OCEAN values.

The five dimensions of OCEAN are Openness (“the active seeking and appreciation of new experiences”); Conscientiousness (“degree of organization, persistence, control, and motivation in goal-directed behavior”); Extraversion (“quan-

¹ Social space is related to 3.6 meters [27].

Fig. 3 Group detection example: six groups detected, being five permanent, in red, and one temporary, in blue



tity and intensity of energy directed outwards in the social world”); Agreeableness (“the kinds of interaction an individual prefers from compassion to tough-mindedness”); and Neuroticism (how much prone to psychological distress the individual is). To detect the OCEAN of each pedestrian, we used the NEO PI-R [15] that is the standard questionnaire measure of the Five-Factor Model. We firstly selected NEO PI-R items related to individual-level crowd characteristics and the corresponding OCEAN-factor. For example: “Like being part of the crowd at sporting events” corresponding to the factor “Extroversion.”

As we describe in detail in [20], we proposed a series of empirically defined equations to map pedestrian features to OCEAN dimensions. Firstly, we 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, we propose 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 an orderly way,” we consider that the individual i should have a high velocity s and low angular variation α to have answer compatible with 5. So the equation for this item is $Q_1 = s_i + \frac{1}{\alpha_i}$. In this way, we empirically defined equations for all 25 items, as presented in [20].

Related to emotion analysis, as we presented in [22], we proposed a way to map the OCEAN dimensions of each pedestrian in OCC Emotion model. This mapping is described in Table 1. In Table 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

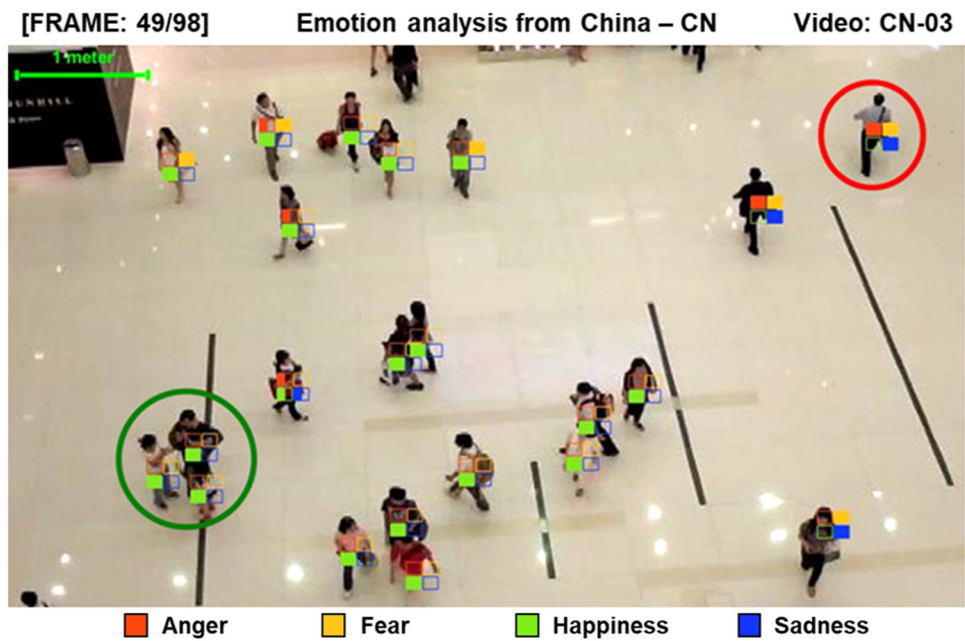
Table 1 Emotion mapping: from OCEAN to OCC

Factor	Fear	Happiness	Sadness	Anger
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

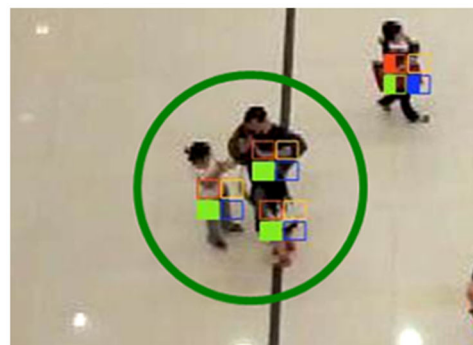
too. A negative value (i.e., -1) does the opposite, therefore, the stronger the factor’s value, the weaker is the given emotion. A zero value means that a given emotion is not affected at all by the given factor. To better illustrate, a hypothetical example is given: if an individual has a high value for Extraversion (for example, $E = 0.9$), following the mapping in Table 1, this individual can present signals of happiness (i.e., If E+ then Happiness= 1) and should not be angry (i.e., If E+ then Anger= -1).

To exemplify, in Fig. 4(a), we highlight two different situations regarding emotions detected using the *Big4GD* model: a group of pedestrians (green circle) and an individual alone (red circle). It is interesting to notice that individuals who are part of a bigger group or have a high collectivity tend to be happy, as we can see in the highlighted group in Fig. 4(b). On the other hand, individuals who are alone and distant from others tend to experience negative emotions (see an example in Fig. 4(c)).

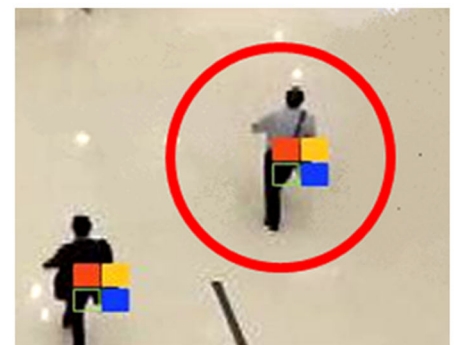
Fig. 4 Emotion analysis in a video from China: the emotion of each individual in scene **a**, focused on a group **b** and one individual **c** highlighted. Each square represents one emotion, where a filled square represents that the person has a positive value from an emotion, a half-filled square means that the emotion is neutral and a not-filled square means that the person has a negative value from an emotion. Each color is related to one emotion: a red square is related to *Anger*, a yellow square means *Fear*, green indicates *Happiness*, and blue, *Sadness*. Based on our approach, individuals in the group **b** tend to be happy, while the individual alone **c** tends to experience negative emotions



(a) Emotion of each individual in the scene



(b) Group



(c) Individual

4.3 Fourth dimension: detection of cultural aspects in crowds

In this section is presented the approach to detect cultural aspects from videos considering Hofstede Cultural Dimensions [32]. In our model, the Hofstede cultural dimensions are related to the *IV - Cultural* geometrical dimension. In order to map pedestrian features in Cultural dimensions, we proposed an approach based on group characteristics [19]. Indeed, collectivism (*COL*) is a % of people grouped, while individualism (*IDV*) is a % of lonely people. Regarding *PDI*, our hypothesis is that individuals that keep close to each other recognize less the group hierarchy, while higher distances between agents can represent a more explicit hierarchy recognition. Hence, we used the mean group distance to describe these cultural dimensions (\bar{d}_g). In terms of *LTO/STO*, the underlying idea is persistence (long-term) as opposed to quick results (short-term). So, we adapted the group orien-

tation to this dimension, meaning that groups with higher values of angular variation result in short-term orientation ($STO = 100 - LTO$), which are computed as shown in Equation 2.

$$LTO = \begin{cases} O_k, & \text{if } O_k \geq 50 \\ 100 - O_k, & \text{otherwise} \end{cases} \quad (2)$$

Considering the *MAS* dimension, we regard that group cohesion can represent “a preference for cooperation.” Thus, higher levels of cohesion represent more feminine values in such dimensions. Indeed, we used also *LTO* to weight the *MAS* aspect: $MAS = \sigma_1 GC_k + (1 - \sigma_1)LTO$, where $\sigma_1 = 0.5$ is the empirically chosen weight. Finally, the *Indulgence vs. restraint* dimension has been characterized by the groups speed and collectivism, given by $IND = \rho_1 S_k + (1 - \rho_1)COL$, where $\rho_1 = 0.5$ is an empirically chosen weight.

Fig. 5 Cultural comparison among the three countries analyzed: Brazil, China, and Austria. We highlight the *IDV* which is lower in Brazil and *LTO* which is higher in Austria. A complete and detailed analysis can be found in [19]

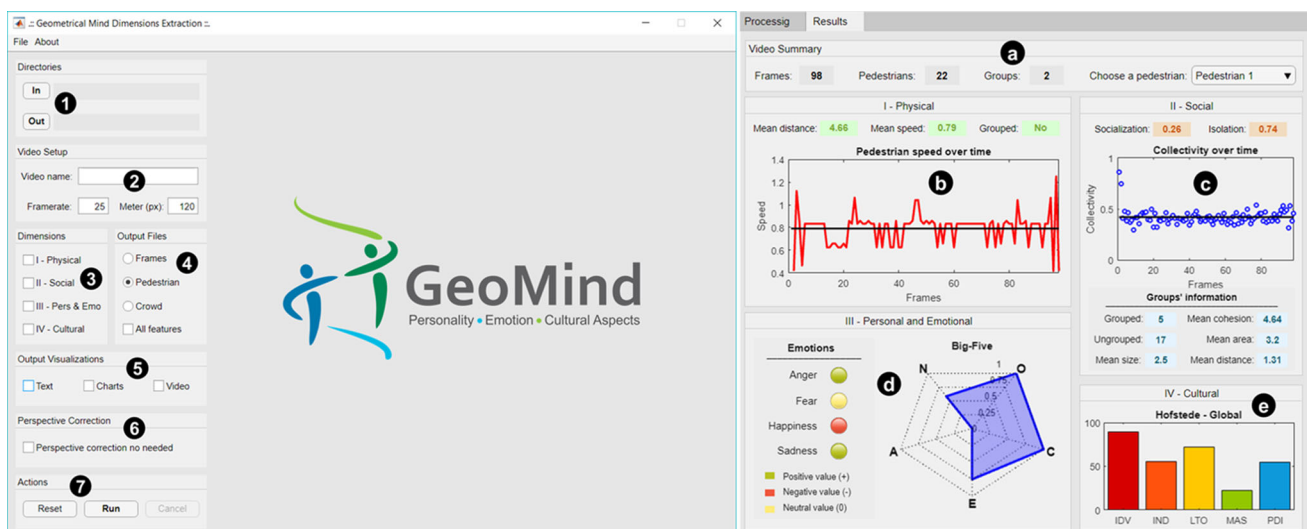
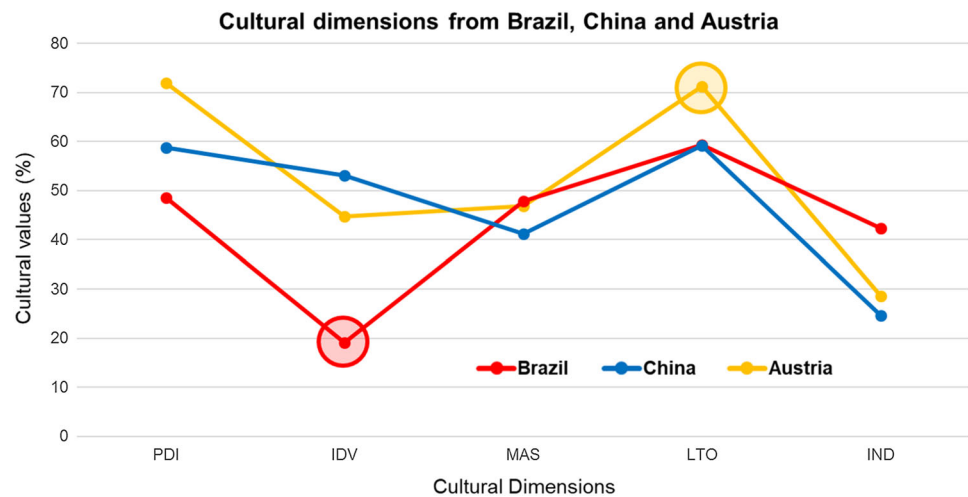


Fig. 6 Main interface of *GeoMind*: setup panel (left) and a video summary after the software processing (right), showing data from “Pedestrian 1”

To exemplify, we show an analysis involving three Countries (Brazil, China, and Austria) which were culturally compared. Figure 5 illustrates the average of each cultural dimension of the three countries of the experiment, with a highlight for the lowest *IDV* (Brazil), as the opposite of China, and the largest *LTO* (Austria).

In the next section, we present the software *GeoMind*, a software developed to extract and analyze pedestrian information from video sequences.

5 Software *GeoMind*

The software, called *GeoMind*² (abbreviated form of Geometrical Mind), was developed using MATLAB App Designer. It was designed to be simple and easy to use, allowing users,

² Download and more information about how to use *GeoMind* can be found at <https://www.rmfavaretto.pro.br/geomind>.

with a few steps, to obtain a series of pedestrian features from video sequences based on tracking. Figure 6 shows the main user interface of the software.

It is possible to see in the left side of Fig. 6 the setup panel, containing the input and output configurations of the video processing:

- 1. Directories:** The input directory must have the video frames (pictures) and the tracking file. The output files generated by *GeoMind* will be stored in the output directory;
- 2. Video Setup:** Users must specify a video name, its framerate, and how many pixels represent one meter in the video;
- 3. Dimensions:** The features were grouped into four dimensions, accordingly to its nature: *I - Physical* with geometrical features, such as speed and angular variation; *II - Social* with features related to groups and social

behaviors, as collectivity and socialization; *III - Personal and Emotional* concerning OCEAN personality and OCC emotion models; and *IV - Cultural* for Hofstede Cultural Dimensions. The users have to select at least one dimension to be computed and save in the output directory;

4. Output Files: In this section, users have to select the frequency in which the features are saved in the output files. Also, they can check the option “All features,” outputting a file with all available information about pedestrians;

5. Output Visualizations: Users can choose how they want to generate information about pedestrians, as text in *.txt* files, as a plot in charts or a video;

6. Perspective Correction: Optionally, users can use a version of the tracking with perspective correction, reducing errors during the calculation of pedestrian positions in the video. If this correction is not necessary, just check the “Perspective correction no needed” option;

7. Action Buttons: Once the user filled all the fields in the setup panel, he has 3 options to choose from: *Run* to start the video processing; *Reset* to restore default values for each predefined form field; and *Cancel*, to stop the processing in progress.

On the right side of Fig. 6, we show the video summary after the software processing. This summary is divided into five areas. Area **a** shows the video information, such as the number of frames, number of pedestrians, and number of groups. In this area, any individual can be selected to see its features (“Pedestrian 1” is selected).

Area **b** shows a summary of the *Physical* features, presenting the mean distance of selected pedestrian to the others, mean speed, an indication if the pedestrian is part of a group or not, and a plot of its speed over time. In area **c**, these are the features related to the *Social* dimension. It shows data about isolation and socialization levels and a plot on the collectivity of the selected pedestrian over time. This section also brings information about groups found in the video, such as the number of grouped and ungrouped pedestrians, mean size of the groups (number of pedestrians), mean cohesion, mean area, and mean distance between the pedestrians in the group.

Area **d** shows the *Personal and Emotion* dimensions features, plotting the emotion values and personalities (Big-Five) of the selected pedestrian in the interval $[0, 1]$. Finally, area **e** is responsible for the *Cultural* features, plotting the Hofstede Cultural Dimensions of the video. The next section presents some experiments and evaluations using the proposed approach.

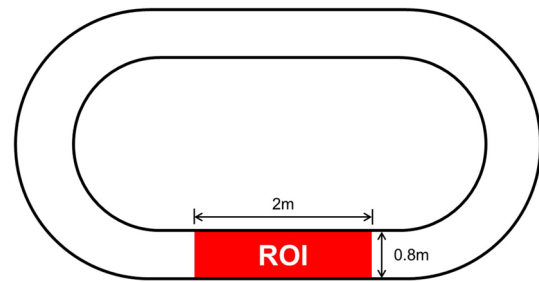


Fig. 7 Sketch of the experimental setup used as proposed by [14]

6 Performed analysis

In this section, we present three analyses we performed regarding geometric features: i) Fundamental Diagram experiment, ii) Personal Space analysis, and iii) a case-study using GeoMind to detect events in video sequences. Each of those analyzes is presented in the following sections.

6.1 Fundamental diagram experiment

We propose a methodology based on three main modules: trajectory detection, data extraction, and FD analysis. For the first component, we obtain the individual trajectories of observed pedestrians from real videos using a tracker [6] where a Fundamental diagram experiment is performed as illustrated in Fig. 7.

This experiment was applied as described in [14] to three countries (India, Germany, and Brazil) with the same populations ($N=15, 20, 25, 30,$ and 34). A corridor was built with markers and tape placed on the ground. Figure 7 presents the size and shape of the corridor. The base of the corridor has a rectangle designating the Region of Interest (ROI) used to capture the population’s data, as proposed in [14].

The experiments conducted in the three countries differed in terms of the moving directions and compositions of the test subjects. For the German study, the group of test persons included male and female students. For the Indian study, the group consisted of only male participants. Participants of the Brazilian study were similar to those of the German study and included male and female students. As presented in the paper by Chattaraj and colleagues [14], the authors performed an analysis regarding the participant selection. The conclusion was that the (mixed) population in Germany and the (male) population in India have free flow speeds which are not statistically different. That indicates that left to themselves both Germans (male and female) and Indians (male) walk similarly in this experimental setup. In our case, based on that, we choose the same setup from the German experiment, with male and female subjects.



Fig. 8 FD experiment performed with $N = 30$ pedestrians: On the left: India [14], in the centre: Germany [14] and on the right: Brazil

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

For the experiment performed in Brazil, we created videos of the same population occupying the same environmental setup as that used for the other countries [14] employing different camera positioning. The camera was positioned overtop of the participants to eliminate the video perspective while for the German and Indian experiments the camera was positioned on the ground in front of the ROI. We used this approach to obtain pedestrian information through computational tracking while in other studies information was collected by a person as pedestrians crossed a demarcated position on the video screen. Figure 8 shows the experiment performed in the three countries with $N = 30$ (where N is the number of pedestrians).

For the Brazilian experiment, as well as in the other experiments, we initially uniformly distributed all individuals throughout the corridor. After instruction was delivered, every individual was instructed to walk around the corridor twice and then leave the environment while continuing to walk a reasonable distance, eliminating the tailback effect. For the second component of our method, we obtained and analyzed the geometric information from trajectories to find neighboring individuals and to compute distances among them. The third module involved fundamental diagram analysis.

6.1.1 Methods

Our research was carried out following relevant international and Brazilian guidelines and regulations. Research protocols used were approved by the Scientific Review Board of the Pontifical Catholic University of Rio Grande do Sul Graduate Program in Computer Science. Informed consent was obtained to use images that could lead to the identification of a study participant and to publish corresponding information/images in an online open-access publication. The survey

was conducted in line with international ethics requirements. All individuals gave written informed consent to participate in this study and they cannot be recognized.

Trajectory Recovery

In the case of India and Germany, we do not have all videos (or tracking) generated from the experiments performed in these countries. Upon request, the authors [14] sent us part of the dataset, i.e., the time (in seconds) at which each individual i enters (t_i^{in}) and exits (t_i^{out}) the experiment ROI (Region of Interest). Based on this information and on the length of the ROI ($2m$ in length), we compute the speed of individual i as

$$s_i = \frac{2}{t_i^{out} - t_i^{in}}, \quad (3)$$

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

$$d_{i,i-1} = s_i |t_i^{in} - t_{i-1}^{in}|, \quad (4)$$

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

For the Brazilian experiment, we use the same information available from other countries for each individual i (i.e., t_i^{in} and t_i^{out}) using a tracking method. Initial individual detection is performed using the method proposed by Viola and Jones [51]. The boosted classifier based on haar-like features was trained with 4500 views of individuals' heads as positive examples and with 1000 other views used as negative examples (the CoffeBreak and Caviar Head datasets from Tosato et al. [48] were used). This detector performs initial position detection based on the positioning of individuals' heads, which are used as input parameters for the next step: tracking. Once individuals are detected, trajectories are obtained using a method proposed by Bins et al. [6]. This approach to object tracking is based on multiple disjoint patches obtained from the target. The patches are represented parametrically by the mean vector and covariance matrix computed from a set of feature vectors that represent each pixel of the target. Each patch is tracked independently from the Bhattacharyya distance [24], and the displacement of the whole template is

Fig. 9 Snapshot of tracking used for the Brazilian experiment. Trajectories (yellow dots) were obtained with the approach developed by Bins, Dihl and Jung [6]

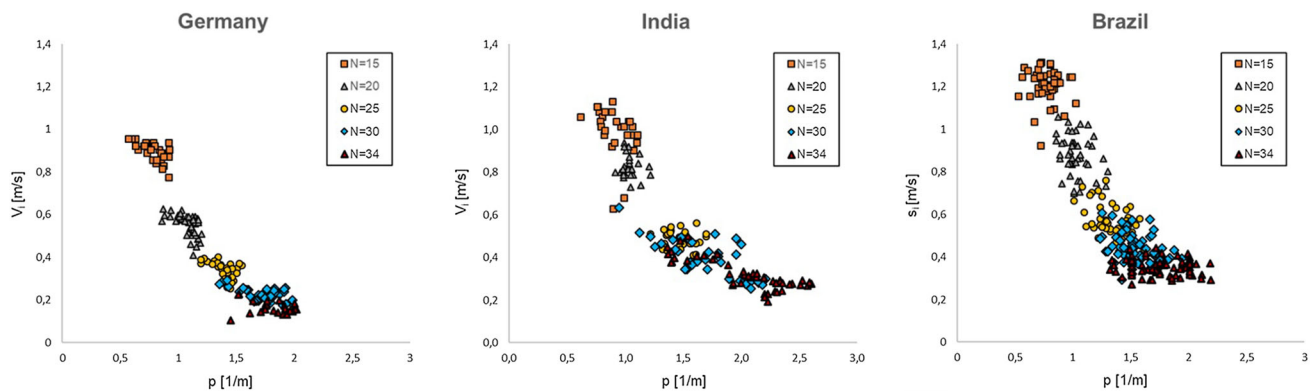
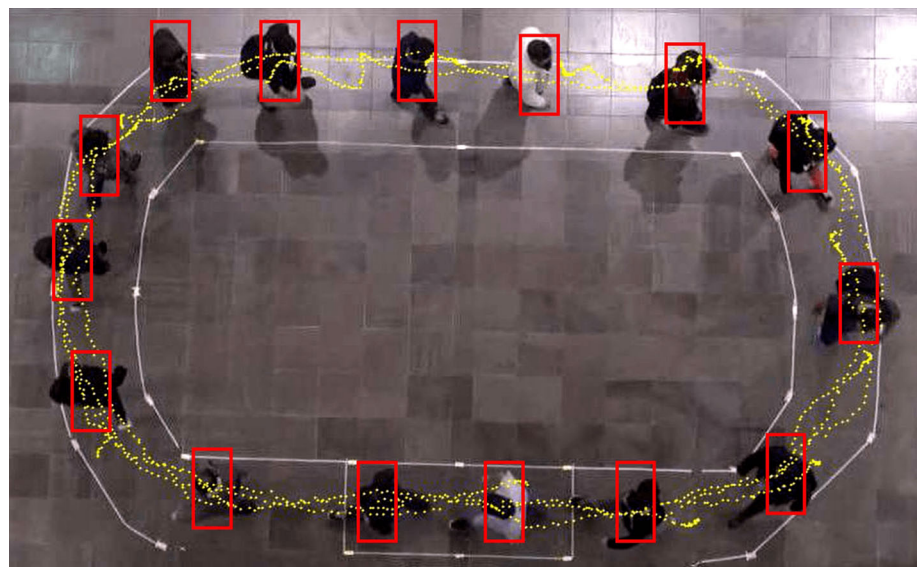


Fig. 10 Speed density data extracted from Fundamental Diagrams for all the three countries: (left) Germany [14], (centre) India [14], and (right) Brazil

obtained using a Weighted Vector Median Filter (WVMF). To smooth the trajectory and to address short-term total occlusions, a predicted displacement vector based on the motion of the target in previous frames is used. Appearance changes of the target are managed by an updating scheme.

As tracker input parameters, we use the initial position of heads detected in the previous step. To obtain the desired parameters for the world coordinate system, we compute the planar homography of each video and transform the extracted trajectories to the world coordinate system by assuming that the head position occupies the ground space ($z = 0$). As our videos are made from a bird's eye view, this assumption does not produce significant errors in projections. The tracker outputs individual trajectories in world coordinates (in Fig. 9 we present a result generated from our work). As noted above, we used information send by authors on FDs for Germany and India (i.e., t_i^{in} and t_i^{out}). For Brazil, we applied a tracking process to extract the 2D position of each individual i at each

frame t : $\mathbf{X}_{i,t}$ (meters). In addition, each time the individual enters in the ROI area we detect t_i^{in} as well as the time when i leaves the area: t_i^{out} . Such information is used in the next step as discussed in the section on *Fundamental Diagram Data Extraction*.

Fundamental Diagram Data Extraction

Such data on pedestrians were used to compute the Fundamental Diagram from the rectangle of analysis. For each frame, we compute the number of individuals within the ROI (2 m-long rectangle of analysis), and the mean population measured over one second (with a video acquisition frame of 30 FPS (frames per second)). Using the same method we computed the mean density and speed of individuals in the ROI, per second. In turn, we determined the density (individuals per sqm) and speed (meters per second) as described by Chattaraj and his collaborators [14] and illustrated in Fig. 10 (on the right). In this figure, we plot data captured from video sequences containing 15, 20, 25, 30, and 34 individuals

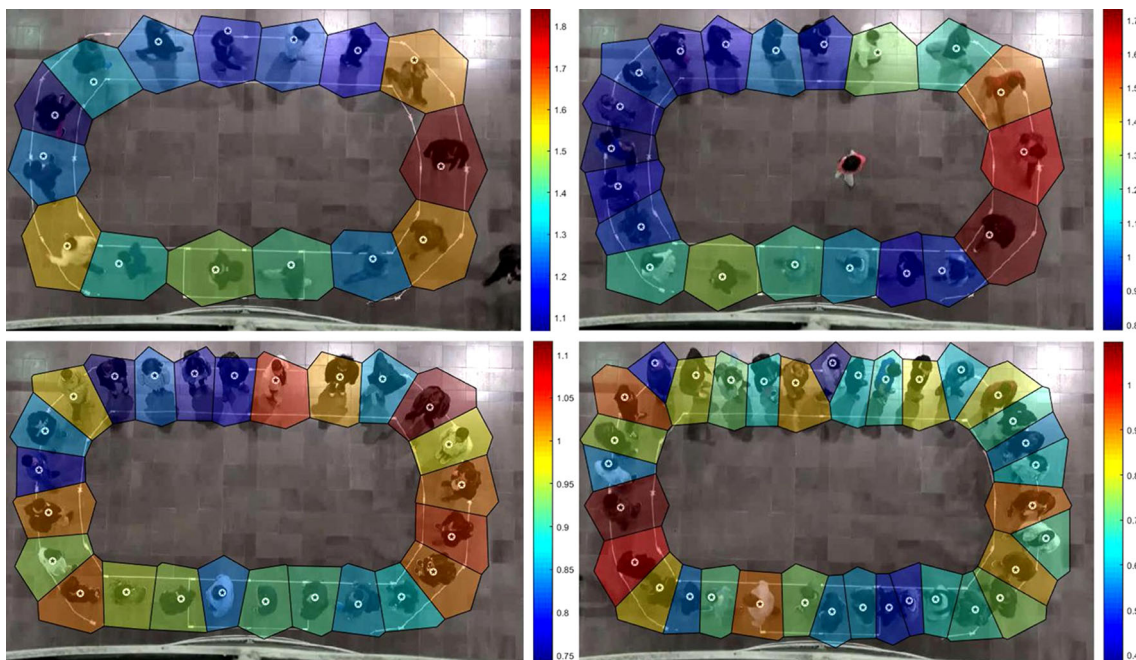


Fig. 11 Voronoi Diagram computed from a video sequence showing, respectively, $N = 15$ and $N = 20$ (top) and $N = 25$ and $N = 34$ (bottom) individuals in Brazil. Values shown on the right denote the personal space (VD polygon area) in metres

while walking in the experimental scenario. As we show, the Brazilian subjects travel at higher speeds than the other two samples.

As computed for the countries Germany and India, the speed for the Brazilian experiment was calculated based on the amount of time it takes for a certain pedestrian i to cross the 2-m ROI, i.e., as described in Equation 3. The mean speed measured at each frame t of the video sequence was computed as follows:

$$\bar{s}_t = \frac{1}{n_t} \sum_{p=1}^{n_t} s_{p,t}, \tag{5}$$

where \bar{s}_t is the mean speed measured at time t . n_t is the number of individuals examined in the rectangle analysis in t , and $s_{p,t}$ is the speed of person p at time t . In a similar way, the density is computed from a 2-m-rectangle analysis as shown in Equation 6:

$$\bar{d}_t = \frac{n_t}{2}. \tag{6}$$

The traditional concept of density (number of individuals by square meter) for the measurement, the region poses a problem due to the small number of individuals involved when analyzing frames. So, we consider the number of individuals found in the region per frame according to the video frame rate (in this case 30 FPS), and we compute the mean number of individuals presented in the ROI per second. In turn, densities are measured as values collected per second.

The following section discusses FD data obtained, i.e., speeds, densities, and personal spaces from the three countries. We adopt the described hypothesis in personal space literature (explaining proxemics as mentioned in Sect. 3) by computing each individual area through Voronoi Diagrams (VD) [2]. As our pedestrian tracking could not be applied to find out the order of pedestrians in the video (we do not know the order in which the pedestrians were tracked, e.g., i and $i + 1$), we use the output of the Voronoi Diagram to compute the neighbor of each individual (pedestrians in front and behind) to calculate distances between each pedestrian and his/her predecessor, which are defined as the personal distance in this work (the magnitude of the vector from individual i to his/her predecessor $i - 1$).

We illustrate personal space (VD polygons) in Fig. 11 for scenarios captured from Brazil ($N = 15$ and $N = 20$ (top) and $N = 25$ and $N = 34$ (bottom)). VD polygons are shown in Fig. 11 provides a visualization of pedestrian distributions observed in the corridor. Small areas (cold colors in the figure) represent the closest pedestrians while larger areas (warm colors in the figure) denote pedestrians that are more distant from one another.

6.1.2 Fundamental diagram analysis

In this section, we discuss some obtained results regarding FD in the three Countries. Figure 12 shows an analysis of the Probability Distribution Function (PDF) applied for the observed personal speeds measured for the three countries.

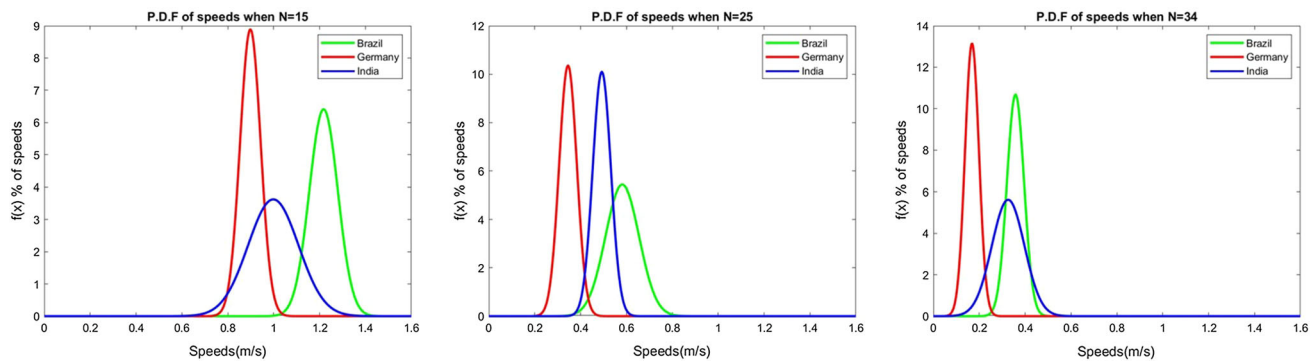


Fig. 12 Probability distribution function (PDF) determined from speeds measured during the experiment: (left) $N = 15$, (centre) $N = 25$, and (right) $N = 34$

Table 2 Mean and standard deviation values of the PDF observed for the countries with 3 N values (15, 25, and 34)

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

PDF function is a statistical expression that defines a probability distribution for a continuous random variable. When the PDF is graphically portrayed, the area under the curve will indicate the interval in which the variable will fall. Figure 12 shows the probability of distributions measured for each observed speed at an interval of $[0; 1.6]$ m/s. Individuals from Brazil present a higher probability of traveling at higher speeds than Indians and Germans. It is also true that as the N increases, the means of PDF curves for the countries tend to converge to a similar value as stated in Table 2 concerning the standard deviation. Thus, although speeds vary more when $N = 34$, the mean values of the different countries are more similar.

Similarly, Figure 13 shows an analysis of the PDF applied to observed personal distances of the three countries for an interval of $[0; 2.5]$ meters. Distances from individuals measured in all countries become more similar as densities increase.

We sought to compare the three countries to find differences/similarities between them. We hypothesize that these differences are extractable through video analysis. Figures 14 and 15 show average speeds and densities as well as corresponding standard deviations for each cluster of the observed population (15, 20, 25, 30, and 34 individuals).

As can be observed in Fig. 14, Brazil presents higher speeds and even when speeds are more heavily impacted, i.e., at higher densities. On the other hand, Germany presents the lowest speeds of the three analyzed countries. It is important to emphasize that the observed standard deviation in speed

and density is higher for Brazil than for the other two countries. When visually observing videos collected from Brazil (videos collected from other countries are not available), one can see that some individuals adopt a “spring” behavior, i.e., sometimes approaching from the person in front and sometimes maintaining distance. Concerning densities achieved by the populations (Figure 15), we note that India presents higher densities than Germany and Brazil. Also, we find that standard deviations for the three countries increase with density as observed for speed.

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

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

1. There is a pattern of FD;
2. As FDs for India, Brazil, and Germany are highly correlated, we assume that the tasks that individuals executed in our experiment were the same, and thus we attempt to measure the cultural factors of this context as discussed in the next section; finally
3. we wish to investigate when the identified differences can be explained by cultural aspects and if they are observable

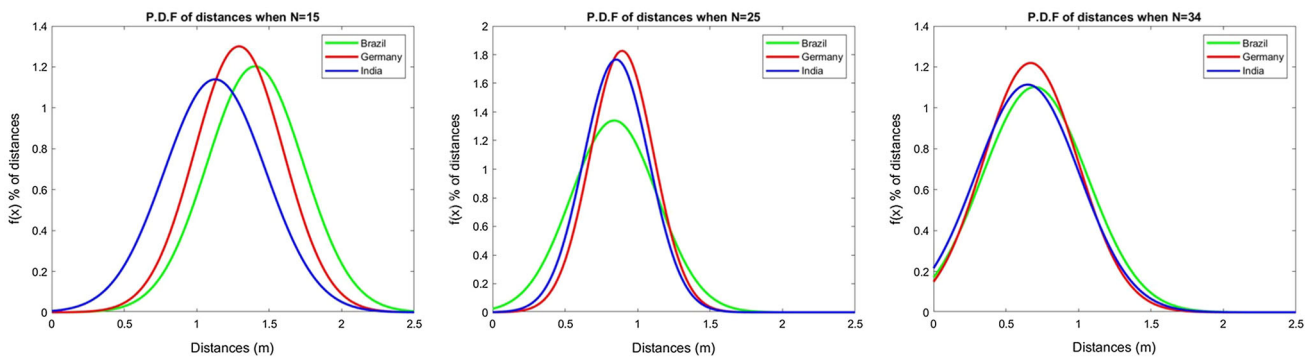


Fig. 13 Probability distribution function (PDF) from the distances in the experiment with, respectively: (left) $N = 15$, (middle) $N = 25$, and (right) $N = 34$

Fig. 14 Walking speed analysis of the three countries and the observed standard deviation

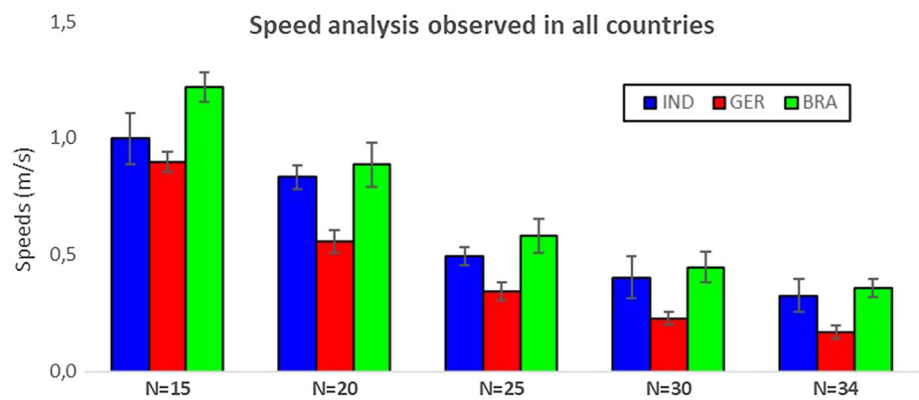
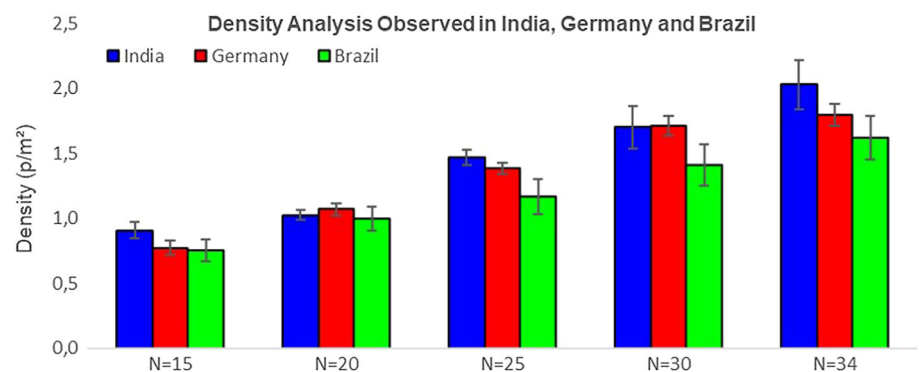


Fig. 15 Density analysis of the three countries and the observed standard deviation



when we investigate personal distance (see the Personal Space Analysis Section below).

6.2 Personal space analysis

In this section, we present our analysis regarding the extracted personal space obtained in the FD experiment. We must first emphasize that as noted in Sect. 3 many studies have been conducted in this area in recent years. In most related works [3,46], the **distance between two individuals is defined as personal space**. We also consider personal space, as we provide an approximation from the Voronoi Diagram (VD), as the VD polygon area. However, for our

analysis, we use the VD to determine the neighborhood of individuals and then to extract those from which to compute distances. In this work, we are interested in investigating cultural aspects of different populations regarding personal distance due to evidence showing that this can differ across countries [46]. As noted in the final section, assuming that the three analyzed populations executed the same task, we extracted personal distances from FDs for our data analysis.

As for the Brazilian population, we had access to all of the experimental video coverage, we were able to track and determine the participants' positions in each frame. However, this was not true for the populations in India and Germany, for which we have only information on the time at which each

person entered and exited the analyzed rectangle as sent by the authors [14]. We thus used this information to compute the speed of each person in the rectangle and the distance between two consecutive individuals in the rectangle (neighbors in the VD). We defined the distance from individual i to the individual in front of him/her (individual $i - 1$) as the personal space of i .

Correlations of distances between the three populations are shown in Fig. 16. As can be easily observed, Pearson's correlations between the populations increase with density. Based on this, our hypothesis that high densities greatly impact personal behavioral expression makes sense as at high densities individuals act more as a mass and less as individuals [49], which ultimately affects their behaviors according to their cultural backgrounds. This assumption is also coherent with one of the main works published on mass behavior [8].

We also evaluated the walking speeds of individuals measured during the FD experiment. Therefore, the corresponding results are very similar to those of the distance analysis as shown in Fig. 17. At lower densities, India and Germany present a higher correlation ($r = 0.42$) while at higher densities all countries present very similar walking speeds. Again, one can speculate that cultural behaviors may be visible at lower and moderate densities, as at higher densities individuals behave like a mass restricted by reduced levels of free space. We must note that this correlation was computed based on all data observed each second and not based on means and standard deviations.

Thus, for this study, we assume that the difference in individual behavior, observed at the same densities and during the execution of the same task, can be considered an expression of individuals' cultural backgrounds. For low to high densities, we observe similarities and differences in terms of personal distances and speeds (Figures 16 and 17). For 15 and 20 individuals measured in the experiments, India and Germany are more similar concerning personal distance and speed than those for Brazil. Indeed, in terms of personal distance, Brazil and Germany present more differentiated behaviors (even inversely proportional as shown by the negative correlation illustrated in Fig. 16) when tested with 15 individuals. Regarding individuals speeds, Brazil vs. India and Brazil vs. Germany pairings present lower correlations for 15 individuals. We hypothesize that this varied behavior observed at low densities can be explained by cultural backgrounds. As stated above, we visually found that Brazilian individuals adopted what we call '*spring behavior*' in the FD experiment, which can be considered as an expression of their social behavior.

Conversely, with increasing density (experiments involving 25, 30, and 34 individuals), obtained personal distances but mainly regarding speeds, higher correlations are found between Brazil and Germany, and an even higher correlation

is found for 34 individuals. In any case, with higher densities, all correlations are higher as well. This may indicate that when space is limited there is less room for individual behaviors to occur. According to our hypothesis, at low densities, India and Germany are more culturally similar. When free space is more limited, it seems that Brazil and Germany were more similar. Finally, at high densities, individuals behave similarly, discarding their cultural behaviors and acting as a mass.

As we show in Fig. 18, regarding the distance vs. speed (on the left) and distance vs. speed exponential relation (on the right) analyses, pedestrians from the different cultures seem to behave similarly at high densities, i.e., over shorter distances (please refer to distances around 0.5, for all countries, the speeds keep around $0.2m/s$). Indian pedestrians, like Brazilians, tend to maintain shorter distances than German pedestrians, and they react less quickly to a change in predecessor speed compared to Germans, as shown in Chattaraj's work ([14]). Brazilian pedestrians seem to act similarly as Indians, but they maintain higher speeds over shorter distances (at higher densities). When we look at low densities, i.e., over higher distances (please refer to the distances around 1.4m) the speeds of all countries are more different.

We also compared preferred distances between individuals identified by Sorokowska and collaborators ([46]) with results obtained from the experiment performed here. Next, we compared data for Germany and Brazil (we were granted access to videos for Germany with populations of 15, 20, 25, and 30 from authors of the PED experimental database³). We must note that the videos from Germany are not the same as those evaluated by Chattaraj [14] and do not involve all N values, e.g., $N = 34$.

In Sorokowska's work, participant answers were given on a distance (0 – 220cm) scale anchored by two human-like figures labeled A and B. Participants were asked to imagine that they were Person A. Participants were then asked to rate how close Person B could approach for them to remain comfortable while having a conversation with Person B. Figure 19 shows a comparison of the four different FD scenarios involving 15, 20, 25, and 30 individuals (as illustrated in the German videos). In our study, we measured the distance person A maintained from person B positioned in front of him or her. For the comparison, from Sorokowska's study, we use an evaluation of acquaintances (individuals who are not close but not strangers) who are comparable to individuals examined in our experiment.

As we show in Fig. 19, while distances generated from our approach are higher than those identified by Sorokowska, proportions identified are similar across all scenarios. People from Brazil maintain a higher distance from others than individuals from Germany. According to our approach, for

³ available at <http://ped.fz-juelich.de/db/>.

Fig. 16 Correlations of personal space measured for the examined countries

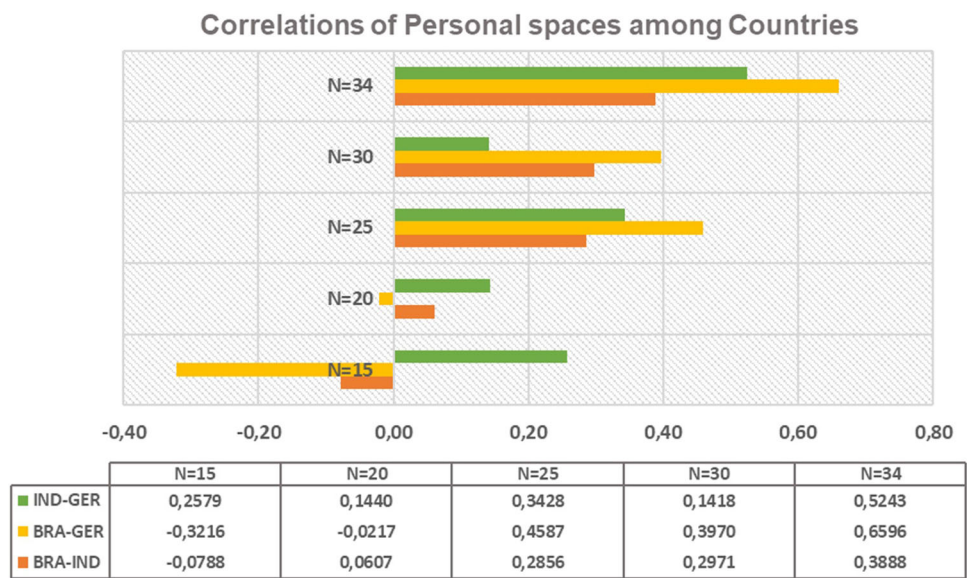


Fig. 17 Correlations of speed measured for the examined populations

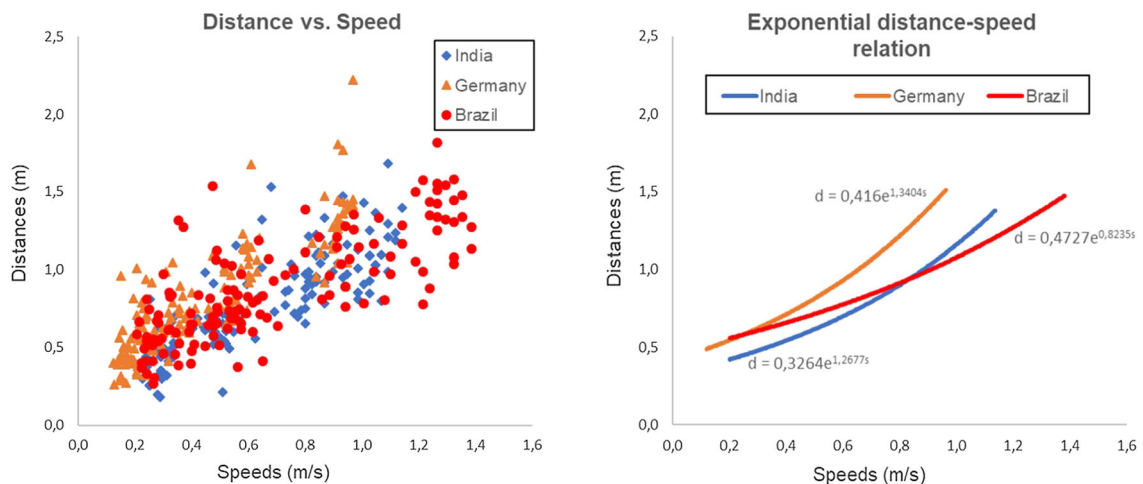
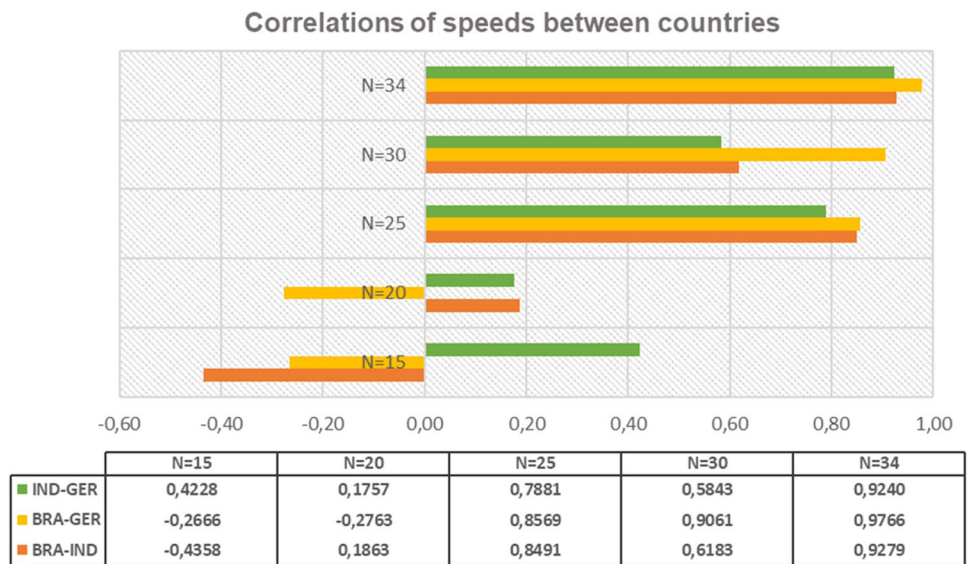


Fig. 18 Distance vs. speed (left) and fitted exponential distance-speed relations (right) analysis with the data from the three countries

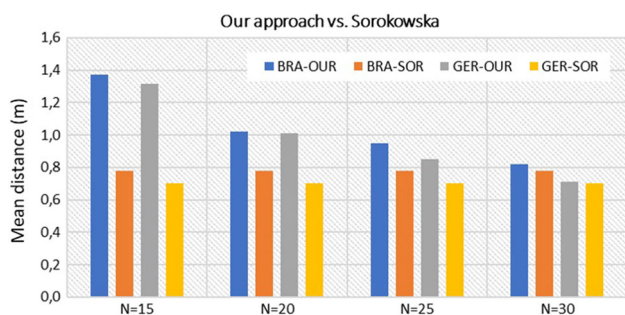


Fig. 19 Our approach versus Sorokowska's [46] for different numbers of individuals examined

the $N = 15$ configuration, individuals from Brazil position themselves 0.5 meters on average further away from one another than they do in Germany while according to Sorokowska individuals from Brazil position themselves 0.8 m further away from one another. It is interesting that as the number of individuals increases, our values become more similar to Sorokowska's (when $N = 30$ the values are quite similar). While the two experiments differ, we show in real conditions individuals behave according to the preferences described in Sorokowska's work. In the next section, we present a practical case-study using *GeoMind*.

6.3 A practical case-study

We performed a case-study using *GeoMind*. We selected three videos that started with a crowd behaving normally (structured behavior) at first and ended with the same crowd exhibiting some kind of unusual behavior (e.g., sudden dispersion, erratic movement, etc.) caused by some event. Those videos were tracked and included as input to *GeoMind*. After processing, *GeoMind* output contains all information specified in the Big-Four Geometric Dimensions, related to each of the pedestrians present in the video at each frame. In the third analysis, we mainly focused on each pedestrian's speed, distance to other pedestrians, and their extracted emotions: anger, sadness, happiness, and fear. Next, we present the evaluation of each selected video.

6.3.1 Video A

Video A, depicted in Fig. 20(a) features footage from a shopping mall in China, taken from the Cultural Crowds database [19]. It serves as our control, representing usual crowd behavior. There are in total 22 pedestrians on the scene, and no anomaly occurs throughout the duration of the video sequence. We are going to analyze the output of 3 different people, each one displaying a distinct behavior. The first one, pedestrian 2 (blue), standing together with a group of people; the second one, pedestrian 6 (orange), walking in a group; and the third one, pedestrian 16 (gray), walking alone.

We can observe that in Fig. 20(b) the speed of each pedestrian stays relatively the same throughout the video. The pedestrians are highlighted in Fig. 20(a) pedestrian 2's (blue) speed remains close to $0m/s$ and pedestrians 6's (orange) and 16's (gray) stays around $0.4m/s$. Figure 20(c) shows the average distance of each pedestrian in relation to every other pedestrian along the sequence. We can see that although the distance is varying, the change is gradual, suggesting that the crowd motion is structured. In Fig. 20(d), we can see the progress of the emotions throughout the frames of the video. As expected, no significant change occurs; barring the initial frames, which are not very representative since the behavior up to that point is unstable.

6.3.2 Video B

The Detection of Unusual Crowd Activity dataset [38] features exclusive footage of people acting out unusual crowd behavior; from this dataset, we have selected Video B, depicted in Fig. 21(a). The video starts with each person moving through the scene either alone or in a group; near the end of the scene, all pedestrians suddenly disperse, as if they're fleeing from some incident. We're going to analyze three of the sixteen people present in the scene: pedestrians 1 (blue) and 2 (orange), walking by themselves; in addition to pedestrian 7 (gray), which is walking side-by-side with someone. In Fig. 21, (a) we have representative frames from before and after the dispersion.

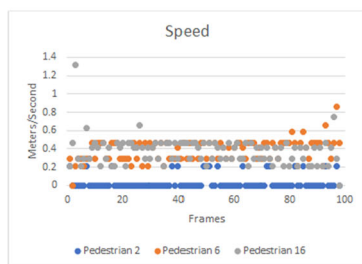
The dispersion occurs at frame 470 and can easily be recognized by observing the speed of the pedestrians, as seen in Fig. 21(b), where the average speed increases by about $1m/s$. The event can also be identified from the distance relative to other pedestrians (c); while it was generally changing gradually before the event, it quickly spikes from an average of $1.47m$ at frame 450 to about $1.95m$ in frame 520. In Fig. 21(d) and (e) we can observe the emotional data of pedestrians 1 and 2, respectively. In both charts, it is possible to see the effects of the event, with fear and sadness arise, whilst happiness decreases; a behavior that is coherent with the situation.

6.3.3 Video C

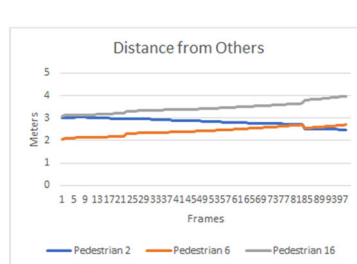
Video C, illustrated in Fig. 22(a), belongs to the BEHAVE Interactions Test Case Scenarios dataset [7], which contains several scenarios of people acting out various interpersonal interactions. We have selected a video of a lone stationary person, pedestrian 1 (blue), who is attacked by four other people, including pedestrian 2 (orange) and 4 (gray); the victim of the attack, then, tries to run away but is eventually tackled to the ground, soon after, the actors leave the frame and thus the video sequence ends.



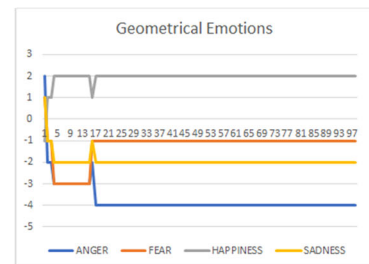
(a) Start and end of Video A, highlighting pedestrian 2 (blue), pedestrian 6 (orange) and pedestrian 16 (gray)



(b) Speed of pedestrians 2, 6, and 16

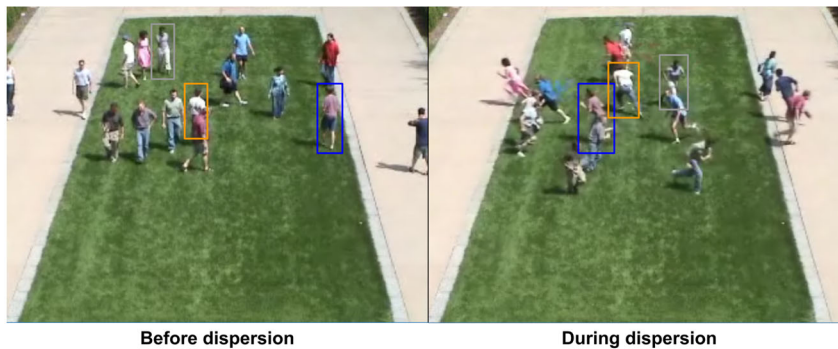


(c) Average distance to others for pedestrians 2, 6, and 16

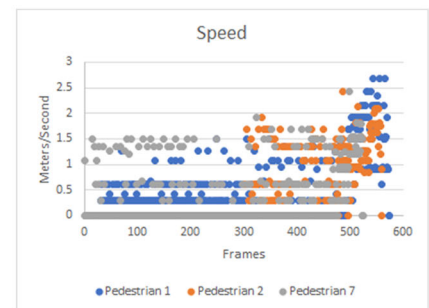


(d) Geometrical emotions for pedestrian 2

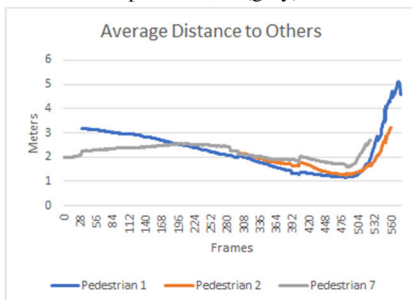
Fig. 20 Video A: Usual crowd activity



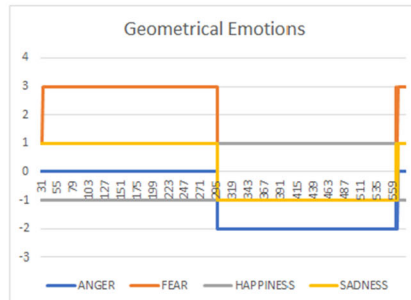
(a) Video B before and during the dispersion event, highlighting pedestrian 1 (blue), pedestrian 2 (orange) and pedestrian 7 (gray)



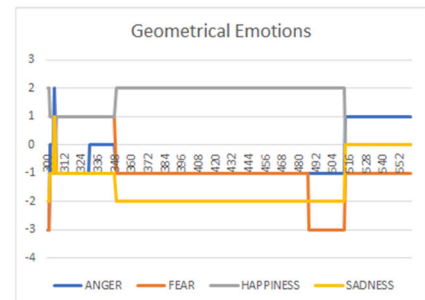
(b) Speed of pedestrians 1, 2, and 7



(c) Average distance to others for pedestrians 1, 2, and 7



(d) Geometrical emotions for pedestrian 1



(e) Geometrical emotions for pedestrian 2

Fig. 21 Video B: Crowd dispersion

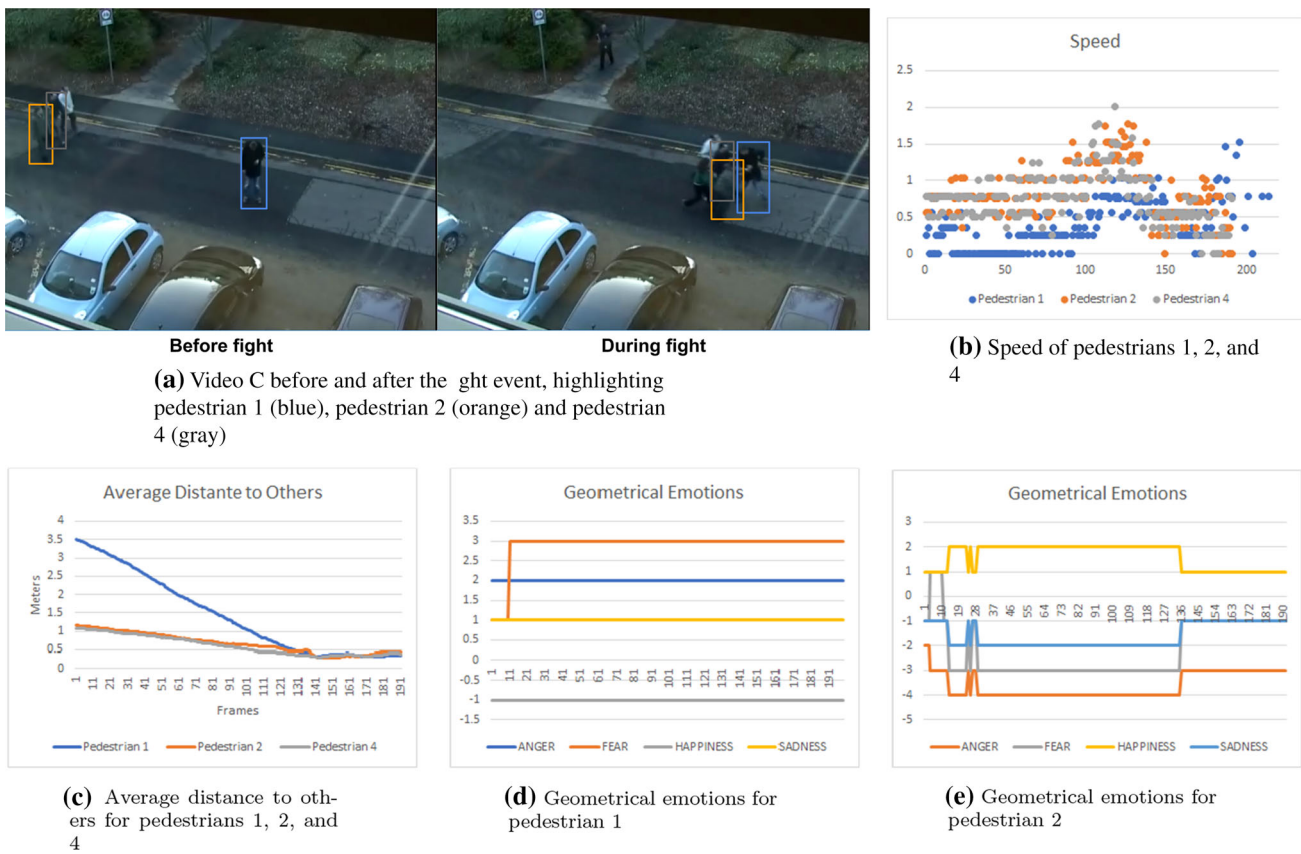


Fig. 22 Video C: Fight activity

In Fig. 22(c), we see the aggressors approaching pedestrian 1 (blue) through its distance to others, which decreases; as expected, because the aggressors are moving in a group, their average distance doesn't change as drastically. Regarding pedestrian speed, the sudden increase in speed as the attackers approach the victim at the start of the fight can be seen clearly around the 100 frame. As the fight progresses and the attackers catch the victim, the speed decreases, as can be seen in the 140 chart. Although these individual components could not be very telling on their own, their composition could be indicative of a fight or similar event. The combination of slowly decreasing distance to others and a spike in velocity could also signal an attack from a group of people in unison.

Regarding the extracted emotions of the pedestrians in the video, pedestrian 1 emotions, as seen in Fig. 22(d), remain somewhat constant throughout the sequence, although the fear component is increased. Pedestrian 2 exhibits the most favorable results regarding the goal to identify events because, as we can see in Fig. 22(e), there is a decrease in happiness and an increase in anger and sadness, which fits the situation at hand. The remaining pedestrians in the video, that is, the other aggressors, do not exhibit any kind of unusual activity.

7 Discussion and final considerations

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

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

We also discussed differences in the cultural features of groups of individuals observed in video sequences. For that, we focused on the *I - Physical* dimension from *Big4GD*. We initially expected to find cross-cultural manifestations to vary from video sequences. However, as one important aspect to be considered in behavior analysis is the context and environment in which individuals behave, we decided to exclude such variations by fixing tasks that the tested populations were required to execute. This is why we used Fundamental Diagrams proposed by [14]. We obtained information for India and Germany and conducted our experiment in Brazil using the same environmental setup. All analysis results presented in this paper are related to these experiments, for which we hypothesized that by fixing the environmental setup and individual tasks, we could evaluate cultural variations in individual behavior.

From this analysis, we make contributions to video understanding and specifically in regard to cultural features of groups of individuals as follows:

- From our FD analysis, we observe decreasing walking speeds with increasing density. This validates FD usage for the three populations. While India and Germany were already evaluated in [14], this paper contributes to data from Brazil. Also, we found higher speeds always observed in Brazil experiments and higher densities observed in India for all tested population sizes (see Figures 14 and 15).
- From our personal space analysis, we found that as the density of individuals increases, individuals are more homogeneous, as shown by the computed Pearson's correlation in given Figures 16 and 17. This shows that in higher densities individuals exhibit mass behaviors instead of behaving individually, according to their cultural background or personalities. This serves as interesting and concrete proof of several theories on mass behavior ([8,49]).
- We think that we have found strong evidence of the fact that the cultural behaviors of individuals can be observed at low densities, and from this, we can define better means to observed events in video sequences. Such cultural behavior is heavily impacted by the speeds at which individuals travel and by distances maintained by others in social spaces. Temporary changes in such behaviors are also important and will be investigated in future work.

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

- Our work shows that in dense crowds, cultural and individual differences are likely to disappear due to limited free space. Thus, the integration of cultural aspects in surveillance systems for measuring pedestrian actions

and behaviors should be performed only at low and moderate densities. In such cases, speeds and distances are important to consider in evaluating cultural aspects.

- When simulating cultural crowds for games or movies as in [33], one should focus on parameter variations (speeds and distances) observed at low and moderate densities of individuals. It can be of benefit to know that at a certain density (high), cultural factors are not as relevant as they are at low and medium densities.
- We must note that when considering safety systems, for example in cases of crowd evacuation, our research can contribute in the sense that in dense crowds, individuals may respond homogeneously, as cultural and individual factors do not seem to interfere in such situations. However, at low and moderate densities, individual differences can produce different results. Of course, in the specific case of crowd behaviors, other variables are also important (e.g., individual training, previous knowledge of the environment, group behavior). We find this to be an interesting topic of research that should be explored to consider the cultural features of crowd evacuation.

We also developed a software called *GeoMind*, which serves as a tool to detect and analyze geometric dimensions in terms of personality, emotion, cultural aspects among other features about pedestrians and groups of people in the video scene. The software was tested with the Cultural Crowds dataset (available at *GeoMind's* website⁴) and results seem promising. This process of extracting information from pedestrians is very useful to better understand people's behavior, allowing analysis, comparisons, and even the use of output files in other applications as games and simulations.

We also performed a practical case-study using *GeoMind* regarding an investigation about events detection in video sequences. Our hypothesis is that we can use individual extracted information to detect events without the training phase, once pedestrian data can include personalities and emotions described using *GeoMind*. Our results could also be used to validate and feed crowd data for simulated scenarios, such as in digital games, where virtual agents must react realistically to player interaction.

Our work has some limitations with respect to the evaluation of extracted data. It is very difficult (or maybe impossible) to assess the pedestrians' emotions, personality, or feelings in order to verify the accuracy of generated data, in video sequences from the internet. That is why we propose this first step in evaluating using Fundamental Diagram, comparisons using Sorokowska [46] and the detection of the events. In future work, we intend to continue to investigate cultural aspects from video sequences measured at moderate

⁴ Available at <https://www.rmfavaretto.pro.br/geomind>.

and low densities to measure the main differences between individuals regarding their behaviors. We will focus on validating speeds and distances as strong indications of cultural features as well as examining other indicators.

As possible future applications, we plan to apply our assumptions to crowd simulations and to improve the accuracy of virtual agents' behaviors based on their personalities and cultural experiences. Also, one of our main goals is to contribute to the event detection field (a widely investigated subject) by identifying unusual and abnormal behavior. We intend to identify a way of differentiating culturally valid behaviors observed in small and medium-sized groups from abnormal behaviors that can have major real-life consequences (e.g., terrorism).

Acknowledgements The authors would like to thank the Office of Naval Research Global (USA) and Brazilian agencies CAPES, CNPQ, and FAPERGS for providing grant support. We also would like to thank Matheus Schreiner Homrich da Silva and Paulo Brossard de Souza Pinto Neto for their contributions in the case-study experiments.

Declarations

Informed consent Research protocols were approved by the Scientific Review Board of the Pontifical Catholic University of Rio Grande do Sul Graduate Program in Computer Science. Informed consent was obtained to use images that could lead to the identification of a study participant and to publish corresponding information/images in an online open-access publication. The survey was conducted in line with international ethics requirements. All individuals gave written informed consent to participate in this study and they cannot be recognized.

Data Accessibility Data could be requested with the author by correspondence.

Conflict of interest The authors declare that they have no conflict of interest.

References

- Aiello, J.R., Thompson, D.E.: *Personal Space, Crowding, and Spatial Behavior in a Cultural Context*, pp. 107–178. Springer US, Boston, MA (1980)
- Aurenhammer, F.: Voronoi diagrams—a survey of a fundamental geometric data structure. *ACM Comput. Surv.* **23**(3), 345–405 (1991)
- Ballendat, T., Marquardt, N., Greenberg, S.: Proxemic interaction: Designing for a proximity and orientation-aware environment. In: *ACM International Conference on Interactive Tabletops and Surfaces. ITS '10*, pp. 121–130. ACM, New York, NY, USA (2010)
- Baxter, J.C.: Interpersonal spacing in natural settings. *Sociometry* **33**(4), 444–456 (1970)
- Bins, J., Dihl, L.L., Jung, C.R.: Target tracking using multiple patches and weighted vector median filters. *J. Math. Imaging Vis.* **45**, 293–307 (2013)
- Bins, J., Dihl, L.L., Jung, C.R.: Target tracking using multiple patches and weighted vector median filters. *MIV* **45**(3), 293–307 (2013)
- Blunsden, S., Fisher, R.: The behave video dataset: ground truthed video for multi-person behavior classification. *Ann. BMVA* **4**(1–12), 4 (2010)
- Bon, G.: *The Crowd: A Study of the Popular Mind*. Cosimo classics personal development, Cosimo Classics (2006)
- Cao, S., Lian, L., Chen, M., Yao, M., Song, W., Fang, Z.: Investigation of difference of fundamental diagrams in pedestrian flow. *Physica A: Stat. Mech. Appl.* **506**, 661–670 (2018)
- Cao, S., Seyfried, A., Zhang, J., Holl, S., Song, W.: Fundamental diagrams for multidirectional pedestrian flows. *JSMTE* **2017**(3), 033404 (2017)
- Castellano, G., Castiello, C., Mencar, C., Vessio, G.: Crowd counting from unmanned aerial vehicles with fully-convolutional neural networks. In: *2020 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8 (2020). <https://doi.org/10.1109/IJCNN48605.2020.9206974>
- Castellano, G., Castiello, C., Mencar, C., Vessio, G.: Crowd detection in aerial images using spatial graphs and fully-convolutional neural networks. *IEEE Access* **8**, 64534–64544 (2020). <https://doi.org/10.1109/ACCESS.2020.2984768>
- Jelić, A., Appert-Rolland, C., Lemerrier, S., Pettré, J.: Properties of pedestrians walking in line: Fundamental diagrams. *Phys. Rev. E* **85**, 036111 (2012)
- Chattaraj, U., Seyfried, A., Chakroborty, P.: Comparison of pedestrian fundamental diagram across cultures. *Adv. Complex Syst.* **12**(03), 393–405 (2009)
- Costa, P., McCrae, R.: *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI)*. Psychological Assessment Resources, New York, NY (1992)
- Diener, E., Diener, M., Diener, C.: Factors predicting the subjective well-being of nations. *J. Pers. Soc. Psychol.* **69**, 851 (1995)
- Digman, J.M.: Personality Structure: Emergence of the Five-Factor Model. *Ann. Rev. Psychol.* **41**(1), 417–440 (1990)
- Evans, G.W., Wener, R.E.: Crowding and personal space invasion on the train: Please don't make me sit in the middle. *J. Environ. Psychol.* **27**(1), 90–94 (2007)
- Favaretto, R.M., Dihl, L., Barreto, R., Musse, S.R.: Using group behaviors to detect hofstede cultural dimensions. In: *International Conference on Image Processing (ICIP)*, pp. 2936–2940. IEEE, Phoenix, AZ (2016)
- 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: *Conference on Graphics. Patterns and Images (SIBGRAPI)*, pp. 223–229. IEEE, Niteroi, RJ (2017)
- Favaretto, R.M., Dihl, L.L., Musse, S.R.: Detecting crowd features in video sequences. In: *2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pp. 201–208 (2016). <https://doi.org/10.1109/SIBGRAPI.2016.036>
- Favaretto, R.M., Knob, P., Musse, S.R., Vilanova, F., Costa, A.B.: Detecting personality and emotion traits in crowds from video sequences. *Mach. Vis. Appl.* **30**(5), 999–1012 (2019)
- Flötteröd, G., Lämmel, G.: Bidirectional pedestrian fundamental diagram. *Trans. Res. Part B: Methodol.* **71**, 194–212 (2015)
- Fukunaga, K.: *Introd. Statist. Patt. Recogn.*, 2nd edn. Academic Press Professional Inc, San Diego, CA, USA (1990)
- HALL, E.T.: A system for the notation of proxemic behavior. *Am. Anthropol.* **65**(5), 1003–1026 (1963)
- Hall, E.T., Birdwhistell, R.L., et al., B.B.: Proxemics [and comments and replies]. *Current Anthropology* **9**(2/3), 83–108 (1968)
- Hall, E.T.E.T.: *The hidden dimension*. New York Anchor (1966). Garden City, N.Y. : Doubleday
- Hillier, B.: A theory of the city as object: or, how spatial laws mediate the social construction of urban space. *Urban Design Int.* **7**, 153–179 (2002)
- Hoch, I.: City size effects, trends, and policies. *Science* **193**(4256), 856–863 (1976)

30. Hofstede, G.: Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations. Sage, Thousand Oaks, CA (2001)
31. Hofstede, G.: Dimensionalizing cultures: The hofstede model in context. ScholarWorks@GVSU. Online Readings in Psychology and Culture (2011)
32. Hofstede, G., Hofstede, G.J., Minkov, M.: Cultures and organizations: Software of the mind. McGraw-Hill, New York, NY (1991)
33. Leandro Dihl Estevão Testa, P.K.e.a.: Generating cultural characters based on hofstede dimensions. In: 2017 IEEE VHCIE, pp. 1–6 (2017)
34. Levine, R.V., Norenzayan, A.: The pace of life in 31 countries. *J. Cross-Cult. Psychol.* **30**(2), 178–205 (1999)
35. Lord, W.: Neo Pi-R - A Guide to Interpretation and Feedback in a Work Context, 1st edn. Hogrefe Ltd, (2007)
36. Mazzon, R., Poiesi, F., Cavallaro, A.: Detection and tracking of groups in crowd. In: 2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance, pp. 202–207 (2013). <https://doi.org/10.1109/AVSS.2013.6636640>
37. Migon Favaretto, R., Rosa dos Santos, R., Raupp Musse, S., Vilanova, F., Brandelli Costa, A.: Investigating cultural aspects in the fundamental diagram using convolutional neural networks and virtual agent simulation. *Computer Animation and Virtual Worlds* **30**(3-4), e1899 (2019). <https://doi.org/10.1002/cav.1899>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/cav.1899>. E1899 cav.1899
38. of Minnesota, U.: Unusual crowd activity dataset of university of minnesota. URL <http://mha.cs.umn.edu/Movies/Crowd-Activity-All.avi>
39. Mora-Colque, R.V.H., Cámara-Chávez, G., Schwartz, W.R.: Detection of groups of people in surveillance videos based on spatio-temporal clues. In: Bayro-Corrochano, E., Hancock, E. (eds.) *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, pp. 948–955. Springer International Publishing, Cham (2014)
40. Nas, P.J.M.: Cities full of symbols: A theory of urban space and culture. Amsterdam University Press, Amsterdam, NL (2011)
41. Nordström, H., Laukka, P., Thingujam, N.S., Schubert, E., Elfensbein, H.A.: Emotion appraisal dimensions inferred from vocal expressions are consistent across cultures: a comparison between australia and india. *Royal Soc. Op. Sci.* **4**(11), 170912 (2017). <https://doi.org/10.1098/rsos.170912>
42. Press, C.U.: Meaning of culture. In: *Cambridge English Dictionary*. Cambridge University Press (2015)
43. Sandıkçı, S., Zinger, S., de With, P.H.N.: Detection of human groups in videos. In: Blanc-Talon, J., Kleihorst, R., Philips, W., Popescu, D., Scheunders, P. (eds.) *Adv. Concept. Intell. Vis. Syst.*, pp. 507–518. Springer, Berlin Heidelberg, Berlin, Heidelberg (2011)
44. Seyfried, A., Steffen, B., Klingsch, W., Boltes, M.: The fundamental diagram of pedestrian movement revisited. *J. Statist. Mech.: Theory Exp.* **2005**(10), P10002 (2005)
45. Sommer, R.: *Personal Space: The Behavioral Basis of Design*. Prentice-Hall (1969). URL <https://books.google.com.br/books?id=j2xLjwEACAAJ>
46. Sorokowska, A., Sorokowski, P., Peter Hilpert, K.C., et al.: Preferred interpersonal distances: A global comparison. *J. Cross-Cult. Psychol.* **48**(4), 577–592 (2017)
47. Suvilehto, J.T., Nummenmaa, L., Harada, T., Dunbar, R.I.M., Hari, R., Turner, R., Sadato, N., Kitada, R.: Cross-cultural similarity in relationship-specific social touching. *Proc. Royal Soc. B: Biol. Sci.* **286**(1901), 20190467 (2019). <https://doi.org/10.1098/rspb.2019.0467>
48. Tosato, D., Spera, M., Cristani, M., Murino, V.: Characterizing humans on riemannian manifolds. *IEEE Trans. Patt. Anal. Mach. Intell.* **35**(8), 1972–1984 (2013)
49. Vilanova, F., Beria, F.M., Angelo Brandelli Costa, Koller, S.H.: Deindividuation: From le bon to the social identity model of deindividuation effects. *Cogent Psychology* **4**(1), 1308104 (2017)
50. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 511–518. IEEE Computer Society, Kauai, HI (2001)
51. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: *IEEE CVPR*, vol. 1, pp. I–511–I–518 vol.1 (2001)
52. Walters, M.L., Dautenhahn, K., Te Boekhorst, R., Koay, K.L., Syrdal, D.S., Nehaniv, C.L.: An empirical framework for human-robot proxemics. *Procs of new frontiers in human-robot interaction* (2009)
53. Werner, C.M., Altman, I., Oxley, D.: Temporal aspects of homes. In: *Home environments*, pp. 1–32. Springer (1985)
54. Wu, N.: A new approach for modeling of fundamental diagrams. *Trans. Res. Part A: Policy and Pract.* **36**(10), 867–884 (2002)
55. Zhang, J.: *Pedestrian fundamental diagrams: Comparative analysis of experiments in different geometries*. Dr., Universität Wuppertal, Jülich (2012). Universität Wuppertal, Diss., (2012)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Rodolfo Migon Favaretto is currently a full professor and researcher at Sul-Rio-Grandense Federal Institute of Education, Science, and Technology, IFSUL, since 2015. He got his PhD degree at Pontifical Catholic University of Rio Grande do Sul (Brazil) in 2019, supervised by Professor Musse, working mainly with emotion, personality, and cultural aspects detection in crowds through computer vision.

Victor Flavio de Andrade Araujo is currently a PhD student in Computer Science at the Pontifical Catholic University of Rio Grande do Sul (BR), supervised by Prof. Soraia Musse. His research interests include the study and evaluation of the human perception of virtual humans and crowds. Araujo has a masters degree in Computer Science from the Pontifical Catholic University of Rio Grande do Sul and a graduate in Computer Science from the Tiradentes University (UNIT) in Aracaju in the state of Sergipe (BR).

Felipe Vilanova is currently a PhD student at the Pontifícia Universidade Católica do Rio Grande do Sul. His main research themes are social attitudes and corruption. Angelo Brandelli Costa is currently a Social Psychology Professor of the Graduate Program in Psychology and the Graduate Program in Social Sciences of the Pontifical Catholic University of Rio Grande do Sul (PUCRS) and coordinator of the Prejudice, Vulnerability and Psychosocial Processes Research Group.

Angelo Brandelli Costa is currently a Social Psychology Professor of the Graduate Program in Psychology and the Graduate Program in Social Sciences of the Pontifical Catholic University of Rio Grande do Sul (PUCRS) and coordinator of the Prejudice, Vulnerability and Psychosocial Processes Research Group.

Soraia Raupp Musse is an associate professor at the School of Technology, Pontifical Catholic University of Rio Grande do Sul (Brazil), where she created and still coordinates the Virtual Human Laboratory. Her research interests include crowd simulation and analysis, facial animation, and integration of computer graphics, pattern recognition, and computer vision. Dr. Musse has a PhD in Computer Science from the Ecole Polytechnique Fédérale de Lausanne (Switzerland), where

she was supervised by Prof. Daniel Thalmann. Recently, she spent a year at UPENN (USA) working as a visiting scholar with Prof. Norman Badler. She has supervised more than 40 graduate students and postdocs and published more than 160 papers. She also published three books in crowd simulation with Springer-Verlag.