USING GROUP BEHAVIORS TO DETECT HOFSTEDE CULTURAL DIMENSIONS

Rodolfo M. Favaretto, Leandro Dihl, Rodrigo Barreto, Soraia Raupp Musse

Graduate Course in Computer Science
Pontifical Catholic University of Rio Grande do Sul, Brazil

ABSTRACT

This paper presents a methodology to characterize information about groups of people with the main goal of detecting cultural aspects. Based on tracked pedestrians, groups are detected and characterized. Group information is then used to find out Cultural aspects in videos, based on the Hofstede cultural dimensions theory. The presented work was tested in videos of pedestrian groups recorded in different countries and results seem promising in order to identify cultural aspects in the filmed sequences.

***Index Terms*—** Groups behaviors, Tracking, Cultural dimensions.

1. INTRODUCTION

Groups and crowd analysis is a phenomenon of great interest in a large number of applications. Surveillance, entertainment and social sciences are examples of areas that can benefit from the development of groups analysis. Currently, with the progress on video processing and computing technology systems, it is possible to develop algorithms to detect and identify groups and to compute crowd features in video sequences. Many works in the literature deal with different applications, for example, counting people in crowds [1, 2], abnormal behavior detection [3, 4], group and crowd movement and formation [5, 6, 7, 8], or detection of social groups in crowds [9, 10, 11, 12]. Most of these approaches are based on personal tracking or optical flow algorithms, and normally consider as features speed, directions and distance over time.

However, there is an important attribute that can influence personal behavior affecting the group that the individual belongs. Chattaraj et al. [13] suggest that cultural and population differences can produce deviations in speed, density and flow of the crowd. In their work, authors discuss the fundamental diagrams used in planning guidelines [14, 15].

In this paper, we propose to use Hofstede's Cultural Dimensions (HCD) theory [16] using information captured from real video sequences. Indeed, HCD is a framework for cross-cultural communication, developed by Geert Hofstede. It describes the effects of a society's culture on the values of its members, and how these values relate to their behaviors. Hofstede executed a large survey study regarding the difference of national values across worldwide subsidiaries of a multinational corporation: he compared the answers of 117,000 IBM matched employees samples on the same attitude survey in different countries. The goal of their research was to find out data about National Culture, which is related with the value differences between groups of nations and/or regions. Hofstede and his collaborators proposed a 6-D model for Dimensions of national cultures: 1) Power distance index (PDI): this dimension is defined as "the extent to which the less powerful members of organizations and institutions accept and expect that power is distributed unequally.". 2) Individualism vs. collectivism (IDV): This index explores the "degree to which people in a society are integrated into groups.". 3) Uncertainty avoidance index (UAI): This index is defined as "a society's tolerance for ambiguity". 4) Masculinity vs. femininity (MAS): In this dimension, femininity is defined as "a preference for cooperation, caring for the weak and quality of life.". 5) Long-term orientation vs. short-term orientation (LTO): a high degree in this index (long-term) views adaptation and circumstantial, pragmatic problem-solving as a necessity. 6) Indulgence vs. restraint (IND): This dimension is essentially a measure of happiness. Indulgent societies believe themselves to be in control of their own life and emotions. Please refer to [17, 18] for more details.

In this paper we propose to detect some of the cultural dimensions defined in HCD [1] using groups behaviors automatically detected in video sequences. For this, we mapped Hofstede's original cultural dimensions to groups characteristics, as described later in this paper. As far as we know this is the first time HCD is used in a computer vision application. The next section discusses the related work, and in Section 3 we present details about the proposed approach, followed by results in Section 4.

2. RELATED WORK

The cultural influence can be regarded in crowds attributes as personal spaces, speed, pedestrian avoidance side and group

Thanks to Office of Naval Research Global (USA) and Brazilian agencies: CAPES, CNPQ and FAPERGS.

[1] We did not consider UAI dimension in our research because we did not find any possible association from images in the database and such dimension.

formations [19], and many works ([20, 21, 5, 12]) focus on the identification of groups using computer vision.

Ge, Collins and Ruback [21] detect small groups of individuals who are walking together. The groups are obtained by bottom-up hierarchical clustering using a generalized, symmetric Hausdorff distance defined with respect to pairwise proximity and velocity. In the work proposed by Chandran, Poh and Vadakkepat [12] a non-recursive motion similarity clustering algorithm is proposed to identify pedestrians traveling together in social groups. People tracking is performed through background subtraction technique using the Mixture of Gaussians approach. Solera et al. [9] propose a new algorithm for the group detection by clustering trajectories and solving it through a parametric correlation clustering trained by a Support Vector Machine (SVM). These works aim mainly to detect social groups from videos of crowds. Others works, in addition to detect the groups, analyze their behaviors [11]. However, groups detection is not trivial in very crowded videos [5, 6], focusing the works only on crowds behavior. Zhou et al. [6] have proposed a descriptor of collectiveness and its efficient computation for the crowd and its constituent individuals. The *Collective Merging* algorithm [6] detects collective motions from random motions.

Our idea is to identify the groups and characterize them to assess the aspects of cultural differences through of the mapping of the Hofstede's dimensions [17, 18]. A similar idea, however using computer simulation and not focused on computer vision, is proposed by Lala et al. [22]. The authors introduced a virtual environment that enables the creation of different types of cultural crowds with which the user may interact. They use Hofstede's dimensions to create a simulated crowd from a cultural perspective.

3. THE PROPOSED APPROACH

Our method presents three main modules: trajectories detection, statistical data extraction and cultural analysis. The first module is able to obtain the individual trajectories of observed pedestrians in real videos (as an alternative, simulated trajectories can be used). In the second module, the statistical information are obtained from trajectories and analyzed in order to find group behavior information. Finally, the last module is responsible for cultural analysis.

3.1. Trajectories Recovery

The people initial detection is performed using the work proposed by Viola and Jones [23]. The boosted classifier working with haar-like features was trained with 4500 views of people heads as positive examples, and 1000 negative (CoffeBreak and Caviar Head datasets from Tosato et al. [24] were used). This detector performs the initial position detection of people based on their heads, which are the input parameters for the next step: tracking. Once the individuals are detected, the trajectories are obtained using a method proposed by Bins et al. [25]. This approach for 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 using the Bhattacharyya distance [26], and the displacement of the whole template is obtained using a Weighted Vector Median Filter (WVMF). To smooth the trajectory and also cope with short-term total occlusions, a predicted displacement vector based on the motion of the target in the previous frames is also used. The appearance changes of the target are handled by an updating scheme.

3.2. Statistical Data Extraction

The main contribution of this paper is to present an approach that, based on features from groups in video sequences, detects cultural aspects. Firstly, we computed information in the individual level for each person i at each timestep: *i)* 2D position x_i of person i, (meters); *ii)* speed v_i of person i (meters/frame); and *iii)* angular variation α_i (degrees) of agent i w.r.t. a reference vector $\vec{r} = (1, 0)$. To obtain the desired parameters in the world coordinate system, we computed the planar homography for each video, and transformed the extracted trajectories to the world coordinate system by assuming that the head position is on the ground place ($z = 0$). Since our videos are close to top-view, this assumption does not produce large errors in the projection. The next section described computed data for groups.

3.2.1. Group Definition and Group Features

Initially, we compute the following parameters for each pair of agents i and j: $s(v_i, v_j)$, $o(\alpha_i, \alpha_j)$ and $d(x_i, x_j)$, where $s(v_i, v_j)$, $o(\alpha_i, \alpha_j)$ are the differences of speed and orientation and $d(x_i, x_j)$ is the Euclidean distance between the two individuals. We use the notion of distances based on the "*proxemics*" described by Hall [27] to define that two agents belong to the same group according with three tests: If ($d(x_i, x_j) <= 1.2$meter) and ($o(\alpha_i, \alpha_j) <= 15°$) and ($s(v_i, v_j) < \beta$), where $\beta = 5\%$ was empirically defined. Based on this rule, agents are grouped in pairs.

In the next step, we check which pairs have one individual in common, and merge them into larger groups. This process is performed until the group formation does not share individuals, i.e. they are disjoint. In the first moment, such established groups are nominated *Temporary Groups*. These groups keep temporary if they stay stable (without inputs and outputs of agents) during at most 10% of total frames of video. After this period, if they keep the group structure, they are classified as *Permanent Groups*. For each *Permanent Group* g we calculated the following data:

- Number of members n_g;

- Mean distances (meters) $\bar{d}_g = 2\frac{\sum_{i=0}^{n_g-1}\sum_{j=0}^{n_g-1} d(x_i,x_j)}{n_g(n_g-1)}$ among all members;
- Mean speeds (meters/frame) $\bar{s}_g = 2\frac{\sum_{i=0}^{n_g-1}\sum_{j=0}^{n_g-1} s(v_i,v_j)}{n_g(n_g-1)}$ among all members;
- Mean angular variation $\bar{\alpha}_g = 2\frac{\sum_{i=0}^{n_g-1}\sum_{j=0}^{n_g-1} o(\alpha_i,\alpha_j)}{n_g(n_g-1)}$ among all members;
- Cohesion of the group C_g, which is described in next section.

In order to have information for each processed video k, we computed the average for all groups from the same video and find final percentages (in order to compare with Hofstede, which values are described in %). For these attributes (except for $\bar{d}$), we computed the percentages (function fp, a simple function to find out the percentage given the maximum value) w.r.t maximum possible obtained values: $S_k = fp(\bar{s}, 1.4)$, $O_k = fp(\bar{\alpha}, 360°)$, $GC_k = fp(\bar{C}, 3)$ (cohesion detailed in next section).

3.2.2. Cohesion

Group cohesion arises when there are strong links among members of a social group. Members of strongly cohesive groups are more inclined to participate readily and to stay with the group [28]. In this paper we propose the analysis of the cohesion of each group inspired on Bassi's work [29]. The amount that an individual wants to remain in its group is quantified through the stability of relationships within groups. More precisely, the group cohesion C_g is given by:

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

where C_i represents the individual cohesion in a certain group g, and recalling that n_g is the number of elements of g. C_i is given by $C_i = CO_i + CV_i + CP_i$ where CO_i is the cohesion due to the individual orientation, CV_i is the cohesion due to the individual speed and CP_i is the cohesion related to the size of the groups. Cohesion factors are calculated by the following empirical equations:

$$CO_i = (w - |\bar{\alpha}_g - \alpha_i|)/w, \tag{2}$$

$$CV_i = (\gamma - |\bar{s}_g - v_i|)/\gamma, \tag{3}$$

$$CP_i = \log n, \tag{4}$$

where $w = 180°$ is related with angle for minimum angular cohesion, $\gamma = 1.4m/s$ is the maximum speed variation related to minimum velocity cohesion. In this work we implemented C_g at each time step for each group in a video sequence. These information are used in the next section in order to map groups information to the cultural dimensions.

3.2.3. Mapping group characteristics for Cultural Dimensions

In order to successfully map groups characteristics in Cultural dimensions, previous research on different knowledge areas was examined. The individualism/collectivism dimension has been considered in [22] as a function of personal space and walking speed. However, in our method we propose to compute this dimension directly via the number of people alone and grouped. Indeed, collectivism (COL) is a % of people grouped, while the individualism (IDV) is a % of lonely people.

The power distance (PDI), masculinity/femininity (MAS), and long/short-term orientation (LTO/STO) have not been addressed in previous researches. In these cases, we try to find out an empirical parametrization to consider these features. In regard to the PDI, which represents the difference between those of high and low status and the behavior towards these groups in society, 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 explicitly hierarchy recognition. Hence, we used the mean group distance to describe these cultural dimension ($\bar{d}_g$). In terms of LTO/STO, the underlying idea is persistence (long-term) as opposed to quick results (short-term). So, we adapted the group orientation 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 5).

$$LTO = \begin{cases} O_k, & \text{if } O_k >= 50 \\ 100 - O_k, & \text{otherwise} \end{cases}. \tag{5}$$

Considering the MAS dimension, we regard that the group cohesion can represent "a preference for cooperation". So, higher levels of cohesion represent more femininity values in such dimension. 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. It is important to notice that we proposed empirical equations to quantify the cultural dimensions in computer vision inspired by HCD, but we are aware that HCD clearly measured the same dimensions but considering a different type of information (subject answers). So, in next section we propose a way to evaluate our work.

4. RESULTS

In this section we discuss some results obtained with our model. Firstly, we evaluated our groups detection in a set of 14 videos (7 from Brazil, 4 from China and 3 from Austria): total of groups in the ground truth is 83 groups, 71 were correctly detected, while false positives= 2 and false negatives=

14. Based on groups behaviors, we present a comparison with Hofstede [18], which provides the cultural dimensions of many countries (`http://geert-hofstede.com/`). We selected 3 of them for which we have some video sequences, in order to process our method and provide a comparison with HCD. We do not want to use HCD as ground truth for our method, since they deal with different information (subject answers in HCD and images in our method); however we want to evaluate if results are coherent to be used to detect cultural aspects. Figure 1 shows the values of cultural dimensions computed in our method and presented in HCD. The obtained average differences for each country, considering all dimensions, are 13.70%, $21,53\%$ and 29.77% for Brazil, China and Austria, respectively. So, we consider that even if we used only a few videos, our results are coherent with HCD. In order to verify if our model allows to detect

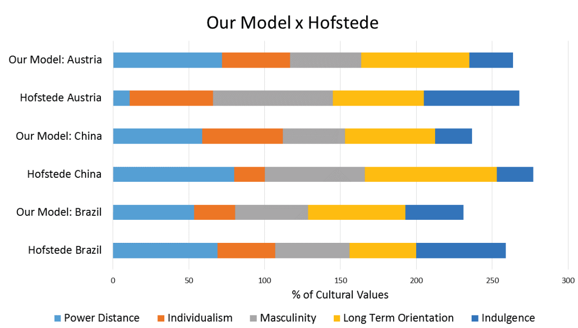

Fig. 1. Values for % of cultural dimensions values.

cultural dimensions from videos, we computed for each of 7 videos from Brazil the average distance in all 5 attributes in comparison with 3 main average groups: Brazil, China and Austria. Results indicate that our method can be used to classify videos from Brazil, based on the cultural dimensions similarity, as illustrated in Figure 2. The two last videos could be misclassified as Austrian videos. Indeed, these two videos presented higher values in both dimensions: PDI (19.84% and 17.55%) and IDV (30.5% and 20.5%) that were greater than the average of Brazilian videos. The third Brazilian video (highlighted in the Figure), obtained the smaller difference between the main group Brazilian. It is illustrated in Figure 3 on the left, and it is easy to perceive the high number of groups. [2] On the other hand, the sixty Brazilian video (highlighted in Figure 3) illustrated more individuals and only a permanent group (blue group in the top of video just appeared in the end of the filmed sequence). In a last analysis, the three Countries (Brazil with 7, China, 4 and Austria with 3 videos) are culturally compared. Figure 4 illustrates the possibilities of our method in order to evaluate the cultural dimensions using our method and studied videos. We highlight the IDV which is lower in Brazil, and LTO that is higher in Austria.

[2]people in the right of the image passed always close to others, so they were never stable to be characterized as a group in our model.

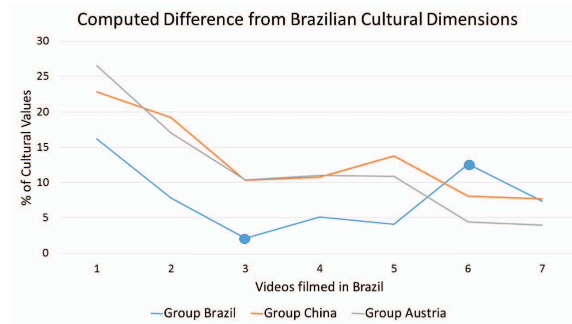

Fig. 2. Computed differences from 7 videos made in Brazil and the three main groups: Brazil, China and Austria.

Fig. 3. Scenes from 3^{rd} and 6^{th} videos from Brazil, respectively. Left: much more groups can be recognized if compared with video on the right.

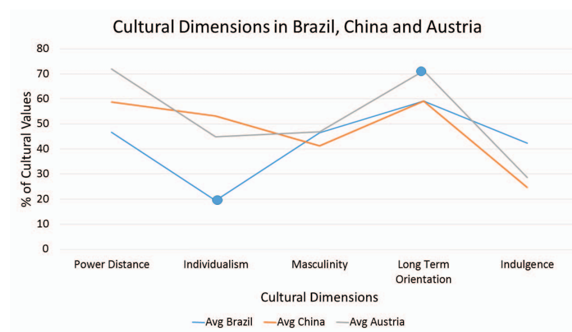

Fig. 4. Cultural Comparison Among the three main studied groups: Brazil, China and Austria.

5. FINAL CONSIDERATIONS

We proposed a methodology to characterize groups of individuals in video sequences and detect some cultural aspects. We were inspired in cultural dimensions proposed by Hofstede. Even with the differences in the used method to capture information (subject answers and computer vision), the results indicate a coherence in our mostly empirical equations. Other information, e.g. crowd size, individual gestures, densities can be used to help detect cultural aspect, which are being addressed in future work.

6. REFERENCES

[1] A.B. Chan and N. Vasconcelos, "Bayesian poisson regression for crowd counting," in *12th IEEE ICCV*, Sept 2009, pp. 545–551.

[2] Zebin Cai, Zhu Liang Yu, Hao Liu, and Ke Zhang, "Counting people in crowded scenes by video analyzing," in *9th IEEE ICIEA*, June 2014, pp. 1841–1845.

[3] E.B. Ermis, V. Saligrama, P. Jodoin, and J. Konrad, "Motion segmentation and abnormal behavior detection via behavior clustering," in *15th IEEE ICIP*, Oct 2008, pp. 769–772.

[4] V. Mahadevan, Weixin Li, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes," in *CVPR 2010*, June 2010, pp. 1975–1981.

[5] B. Solmaz, B. E. Moore, and M. Shah, "Identifying behaviors in crowd scenes using stability analysis for dynamical systems," *IEEE PAMI*, vol. 34, no. 10, pp. 2064–2070, Oct. 2012.

[6] Bolei Zhou, Xiaoou Tang, Hepeng Zhang, and Xiaogang Wang, "Measuring crowd collectiveness," *IEEE PAMI*, vol. 36, no. 8, pp. 1586–1599, Aug 2014.

[7] Ricky J. Sethi, "Towards defining groups and crowds in video using the atomic group actions dataset," in *2015 IEEE International Conference on Image Processing, ICIP 2015, Quebec City, QC, Canada, September 27-30, 2015*, 2015, pp. 2925–2929.

[8] Hyunjoon Jo, Kabir Chug, and Ricky J Sethi, "A review of physics-based methods for group and crowd analysis in computer vision," *Journal of Postdoctoral Research*, vol. 1, no. 1, pp. 4–7, 2013.

[9] Francesco Solera, Simone Calderara, and Rita Cucchiara, "Structured learning for detection of social groups in crowd," in *10th IEEE AVSS*, Aug. 2013.

[10] Jing Shao, C.C. Loy, and Xiaogang Wang, "Scene-independent group profiling in crowd," in *IEEE CVPR*, June 2014, pp. 2227–2234.

[11] Linan Feng and B. Bhanu, "Understanding dynamic social grouping behaviors of pedestrians," *IEEE STSP*, vol. 9, no. 2, pp. 317–329, March 2015.

[12] A.K. Chandran, Loh Ai Poh, and P. Vadakkepat, "Identifying social groups in pedestrian crowd videos," in *ICAPR*, Jan 2015, pp. 1–6.

[13] U. Chattaraj, A. Seyfried, and P. Chakroborty, "Comparison of pedestrian fundamental diagram across cultures," *ACS*, vol. 12, no. 03, pp. 393–405, 2009.

[14] U. Weidmann, "Transporttechnik der fussgnger," Tech. Rep., Institut fr Verkehrsplanung, ETH Zrich,, 1993.

[15] A. I. Predtechenskii V. M., Milinskii, *Planning for foot traffic flow in buildings*, New Delhi : Amerind,, 1978.

[16] G. Hofstede, *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations*, Sage, Thousand Oaks, CA, 2001.

[17] G. Hofstede, *Cultures and organizations: software of the mind.*, London: McGraw-Hill., 1991.

[18] G. Hofstede, "Dimensionalizing cultures: The hofstede model in context," *ScholarWorks@GVSU. Online Readings in Psychology and Culture*.

[19] N. Fridman, G. A. Kaminka, and A. Zilka, "The impact of culture on crowd dynamics: An empirical approach," in *AAMAS*, 2013, pp. 143–150.

[20] B. Zhan, D. Monekosso, P. Remagnino, S. A. Velastin, and L. Xu, "Crowd analysis: a survey," *MVA*, vol. 19, no. 5-6, pp. 345–357, 2008.

[21] Weina G., Robert T. C., and Barry R., "Vision-based analysis of small groups in pedestrian crowds," *IEEE PAMI*, vol. 34, no. 5, pp. 1003–1016, 2012.

[22] D. Lala, S. Thovuttikul, and T. Nishida, "Towards a virtual environment for capturing behavior in cultural crowds," in *6th ICDIM*, Sept 2011, pp. 310–315.

[23] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *IEEE CVPR*, 2001, vol. 1, pp. I–511–I–518 vol.1.

[24] D. Tosato, M. Spera, M. Cristani, and V. Murino, "Characterizing humans on riemannian manifolds," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1972–1984, 2013.

[25] J. Bins, L. L. Dihl, and C. R. Jung, "Target tracking using multiple patches and weighted vector median filters," *MIV*, vol. 45, no. 3, pp. 293–307, Mar. 2013.

[26] K. Fukunaga, *Introduction to statistical pattern recognition (2nd ed.)*, Academic Press Professional, Inc., San Diego, CA, USA, 1990, 1990.

[27] E.T. Hall, *The Hidden Dimension*, A Doubleday anchor book. Anchor Books, 1990.

[28] Dyaram L.and Kamalanabhan T. J., "Unearthed: The other side of group cohesiveness," *Journal of Social Sciences*, vol. 10, no. 3, pp. 185–190, 2005.

[29] Bassi Bruce, "Computer simulation of crowd dynamics and destructive crowd behavior," Honors Scholar Theses. Paper 20., 2006.