Generating Cultural Characters based on Hofstede Dimensions

Leandro Dihl
PUC-RS, Brazil
E-mail: leandro.dihl@acad.pucrs.br

Estêvão S Testa
PUC-RS, Brazil
E-mail: estevao.testa@acad.pucrs.br

Paulo Knob
PUC-RS, Brazil
E-mail: paulo.knob@acad.pucrs.br

Gabriel L. B. da Silva
PUC-RS, Brazil
E-mail: gabriellimab@gmail.com

Rodolfo M. Favaretto
PUC-RS, Brazil
E-mail: rodolfo.favaretto@acad.pucrs.br

Marlon F. de Alcântara
UMass Lowell-MA,USA
E-mail: marlonmfa@gmail.com

Soraia R. Musse
PUC-RS, Brazil
E-mail: soraia.musse@pucrs.br

***Abstract*—The virtual humans behaviors can be endowed with different levels of intelligence and their action/motion can present more or less realism. In this scope, it is still a challenge to create, in an automatic and easy way, virtual humans that move and behave in a way that seems natural. Cultural aspects and population differences can produce deviations in speed, density and flow of the crowd. These aspects can be observed in videos of different Countries and certainly it can produce more realistic agents, if incorporated to simulation models. This paper presents a methodology to generate virtual humans which trajectories are based on real life video sequences, and in addition we considered Hofstede Cultural Dimensions (HCD) to map cultural differences. The method is tested in a set of real environments, but can be scaled for any environment that have video sequences and tracked human trajectories for training..**

I. Introduction

Modern technologies have provided greater realism in computer games/simulation. However, most of the such environments still need to better reflect the real world. In particular, most Non-Playable Characters (NPCs) are limited by a restricted range of trajectories, defined by the game designers. This aspect, relevant in background population, when unrealistic, can make the game boring for the players. Moreover, background characters are typically programmed by game designers in a process that requires many resources, is very time consuming and can be quite tedious. One way to deal with this problem is using data coming from real life in order to guide virtual humans behaviors. This data can be obtained from real videos and crowd analysis concepts.

Indeed, crowd/human 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] and 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].

Using a different approach, Hofstede [16] in his most notable work has developed cultural dimensions theory. He described national cultures along six dimensions: Power Distance, Individualism, Uncertainty avoidance, Masculinity, Long Term Orientation, and Indulgence vs. restraint.

Therefore, the main goal of this paper is to propose a way to generate individuals behaviors based on Cultural dimensions from Hofstede [16] and applied trajectories from real video sequences. It is not the first time computer vision and games are related. As presented by Freeman et al [17] in 1996, ”the appeal of computer games may be enhanced by vision-based user inputs”. Certainly it was the case of last decades with the usage of computer vision in interfaces.

This paper is organized as follows: Section II describes the literature review on the topic of the cultural influence in groups and crowds, the work presented by Hofstede and the virtual humans behaviors. In the Section III, we presents how the features are obtained and mapped to be applied in virtual humans control. The proposed model is presented in Section IV, while experimental results and final considerations are discussed in Section V.

II. Related Work

The cultural influence can be regarded in crowds attributes as personal spaces, speed, pedestrian avoidance side and group formations [18], and many work ([19], [20], [5], [12]) are focused on the identification of groups using computer vision.

Ge, Collins and Ruback [20] 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 methods aim mainly to detect social groups from videos of crowds. Other works, in addition to detecting the groups, also 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 characterize the virtual agents mapping Hofstede's Cultural Dimensions (HCD) theory [16], [21] using information captured from crowds real video sequences, as proposed by Favaretto [22]. Thus, we can model virtual human groups having different characteristics as mapped using Hofstede. A similar idea, however using computer simulation and not focused on computer vision, is proposed by Lala et al. [23]. 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.

III. Generating behaviors based on Hofstede Cultural Dimensions

In this paper, we proposed to use Hofstede's Cultural Dimensions (HCD) theory [24] based on information captured from real video sequences and then to apply captured characteristics in virtual humans behavior. 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 [16], [21] for more details.

Using real crowd videos from different places we obtained a set of features, and use that to characterize the virtual human groups. To obtain these features we applied the methods presented by Favaretto et al [22]. In this work, the authors presented a methodology to characterize information about groups of people with the main goal of detecting cultural aspects. Based on tracked pedestrians, groups were detected and characterized. Group information was then used to find out Cultural aspects in videos, based on the Hofstede cultural dimensions theory. The work was tested in videos of pedestrian groups recorded in different countries and results seemed promising in order to identify cultural aspects in the filmed sequences. Based on the work developed by Favaretto and collaborators, we could calculate the HCD for each video sequence. These values are given by pdi representing the Power distance index, the mas that is the Masculinity vs. femininity, lto that is the Long-term orientation vs. short-term orientation and finally the ind that represent the Indulgence vs. restraint of the groups. The remaining dimensions IDV and UAI were not used in [22], so in this first approach we also discarded such parameters, that certainly must be included in a future work.

Our idea was to use the inverse functions presented by Favaretto et al. [22] to determine the mean distance, mean cohesion, mean angular variation and mean speeds (meters/frame) according to the videos and so applying that features on virtual crowds. Our hypothesis is that we are able to generate coherent virtual human simulation, if compared with a originated video sequence (HCD values).

Once we have the values for pdi, ind, lto and mas extracted from video sequences, we applied the equations proposed in [22] as well as the maximum values, in order to estimate the following data to simulate the groups in the crowd. so, for each group g to be simulated:

- Mean distance (meters) is given by $\bar{d}_g = ((100 - pdi).(1.2)/100)$,
- Mean speed (meters/frame) is given by $\bar{s}_g =$

Fig. 1. The group features are obtained applying the approach presented by Favaretto et al [22]. On the top: the input image and heads detection, and on the bottom: the people detection and tracked trajectories.

$(ind.1.4/100)$,

- Mean angular variation is given by $\bar{\alpha}_g = ((100 - lto)/100)$, and
- Mean cohesion of the group is given by $\bar{\gamma}_g = (((100 - mas) * 3)/100)$.

Based on these simple equations, we generated $\bar{d}_g$, $\bar{s}_g$, $\bar{\alpha}_g$ and $\bar{\gamma}_g$ for each group g to be simulated. Next section shows how we use this data and the trajectories from video sequences to generate virtual humans.

IV. Applying HCD to the virtual humans

We use the method proposed by Alcantara et al. [25] to control the virtual agents based on video sequences. In the case of this work, we included the HCD parameters to control the groups behaviors.

Firstly, the number of agents and groups are created in two possible ways: the first one makes such definition according to the input video (using the method proposed in [25]), i.e. the number of agents and groups corresponds exactly to the data provided in the video sequence and extracted using computer vision. The second possible way is the direct definition based on input parameters (in a input txt file used as script).

Secondly, trajectories from the filmed sequence can be used to define initial and ending positions, as presented in [25] or again it can be provided in the input file to guide the simulation. In the first case, we consider an individual trajectory as a set of positions $\vec{X}_i$ in 2d world with F elements, representing the F frames of motion duration. The set of trajectories in a specific video sequence is used in order to learn the existent patterns. In fact, the individuals trajectories extracted from video sequences are used to defined the set of possible movements the agents will apply in the virtual environment. In the second option, initial and ending positions are interpolated to generate possible trajectories, in this case totally disconnected from the filmed sequence.

Agents and groups are created according to the input information. So, for a specific agent a, associated to a group g, we select a set of of possible moves (MOV_a) considering a chosen specific trajectory Equation 1:

$$MOV_a = \bigcup_{f=1}^{F-1} \{\Delta\vec{X}_i^f\}, \tag{1}$$

where $\Delta\vec{X}_i^f = \vec{X}_i^{f+1} - \vec{X}_i^f$. this represent the basic motion each agent is going to apply. In addition to that, the HCD parameters are taken into account in order to provide parameters for simulations. Specifically for each group g: $\bar{d}_g$, $\bar{s}_g$, $\bar{\alpha}_g$ and $\bar{\gamma}_g$. Next sections detail the process of creation and simulation of agents.

A. Creating Virtual Agents

Firstly, we define the initial agent a attributes as follows:

- Identity (id_a) is the unique value that identifies the agent a;
- Group (a_g) is the group that agents a is part;
- Movement possibilities based on possible trajectories MOV_a (as defined in Equation 1).

Once we have the groups and agents, their attributes and trajectories defined, we used the HCD parameters to guide the behaviors. Firstly, the mean distance $\bar{d}_g$ and mean speed $\bar{s}_g$ should be applied as a function of time for the agents present in group g. It is easily regulated computing the initial speeds in order to achieve the expected mean value and sharing initial positions and goals for each agent into group g. In the same way, the orientation $\bar{\alpha}_g$ is applied randomly among the agents from group g. The standard deviation was empirically chosen from 0% to 10% of maximum values for each attributes.

To represent the Cohesion, as proposed in [22], we included a simplified version of Reynolds Boids [26] to make agents from the same group trying to match their parameters, i.e. the mean speed. Although agents speed are initialized in order to achieve the expected mean distance, during the simulation we can re-compute that to consider the cohesion. Indeed, agents try to reach the mean speed (by reducing the standard deviation) in their individual speed. How much the agents try to achieve the average is proportional to the group cohesion level, so the higher cohesion determines agents trying to have exactly the same speeds.

V. Preliminary Results and Final Considerations

We developed an application using Unity to read a short video and based on the tracked information, extract HCD data as well as its trajectories. The experimental results have essentially two stages, the video analysis and the simulation that generates groups/individuals simulation

As an input to our method, the Cultural Crowds [22] dataset was used. It consists of 33 crowded video clips from different countries. Some videos were manually acquired by the authors while others come from public datasets, like Collective Motion [27] and Data Driven Crowd [28], selected from Getty Images [29].

The Figure 2 illustrates the proposed method, related to the real video sequence UKN-03. In this case, we use the trajectories coming from the real video (Figure 2(a)), and number of agents and groups coming from input script(Figure 2(b)). In the real video, there are 20 agents and 6 groups, mean speed and mean distance among agents from the same groups are $0.88m/s$ and $0.74m$, respectively. The data originated in the simulation are: 88 agents, 24 groups, with mean speed of $0.88m/s$ and mean distance of $0.77m$. Later, a visualization of the agents was generated (Figure 2(c)).

This is an ongoing work and further analysis of cultural results is still in development. For the evaluation, the videos should be quantitatively compared to the respective simulation in terms of HCD parameters, according to [22]. However, the qualitative assessment of results is still on going, since we are planning to measure if the human perception about HCD cultures is valid on virtual humans simulation. This is the future work.

(a)

(b)

(c)

Fig. 2. (a) the input image with people detection, in (b): generated virtual humans positions over the extracted background, and (c): the visualization of the generated humans over the extracted background.

Acknowledgment

The authors wish to thanks the Office of Naval Research Global (USA) and Brazilian agencies: CAPES, CNPQ and FAPERGS.

References

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

[2] Z. Cai, Z. L. Yu, H. Liu, and K. Zhang, "Counting people in crowded scenes by video analyzing," in *9th IEEE ICIEA*, June 2014, pp. 1841–1845.

[3] E. 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, W. 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. [Online]. Available: http://dx.doi.org/10.1109/TPAMI.2012.123

[6] B. Zhou, X. Tang, H. Zhang, and X. Wang, "Measuring crowd collectiveness," *IEEE PAMI*, vol. 36, no. 8, pp. 1586–1599, Aug 2014.

[7] R. 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. [Online]. Available: http://dx.doi.org/10.1109/ICIP.2015.7351338

[8] H. Jo, K. Chug, and R. 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] F. Solera, S. Calderara, and R. Cucchiara, "Structured learning for detection of social groups in crowd," in *10th IEEE AVSS*, Aug. 2013.

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

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

[12] A. Chandran, L. A. 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. [Online]. Available: http://www.worldscientific.com/doi/abs/10.1142/S0219525909002209

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

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

[16] G. Hofstede, *Cultures and organizations: software of the mind.* London: McGraw-Hill., 1991. [Online]. Available: http://books.google.com.br/books?id=zGYPwLj2dCoC

[17] W. T. Freeman, K. Tanaka, J. Ohta, and K. Kyuma, "Computer vision for computer games," in *Automatic Face and Gesture Recognition, 1996., Proceedings of the Second International Conference on*, Oct 1996, pp. 100–105.

[18] N. Fridman, G. A. Kaminka, and A. Zilka, "The impact of culture on crowd dynamics: An empirical approach," in *AAMAS*, 2013, pp. 143–150. [Online]. Available: http://dl.acm.org/citation.cfm?id=2484920.2484946

[19] 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. [Online]. Available: http://dx.doi.org/10.1007/s00138-008-0132-4

[20] W. G., R. T. C., and B. R., "Vision-based analysis of small groups in pedestrian crowds," *IEEE PAMI*, vol. 34, no. 5, pp. 1003–1016, 2012. [Online]. Available: http://doi.ieeecomputersociety.org/10.1109/TPAMI.2011.176

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

[22] R. M. Favaretto, L. Dihl, R. Barreto, and S. R. Musse, "Using group behaviors to detect hofstede cultural dimensions," in *2016 IEEE International Conference on Image Processing (ICIP)*, Sept 2016, pp. 2936–2940.

[23] 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.

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

[25] M. Alcantara, E. Testa, G. Lima, R. Favaretto, L. Dihl, and S. R. Musse, "Generating background population for games based on real video sequences." in *Proceedings of SBgames*, SBGames, Ed., 2016.

[26] C. W. Reynolds, "Flocks, herds and schools: A distributed behavioral model," *SIGGRAPH Comput. Graph.*, vol. 21, no. 4, pp. 25–34, Aug. 1987. [Online]. Available: http://doi.acm.org/10.1145/37402.37406

[27] B. Zhou, X. Tang, and X. Wang, "Measuring crowd collectiveness," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, June 2013, pp. 3049–3056. [Online]. Available: http://dx.doi.org/10.1109/CVPR.2013.392

[28] M. Rodriguez, J. Sivic, I. Laptev, and J.-Y. Audibert, "Data-driven crowd analysis in videos," in *Proceedings of the International Conference on Computer Vision (ICCV)*, 2011. [Online]. Available: http://dx.doi.org/10.1109/ICCV.2011.6126374

[29] J. P. G. Trust, "Getty provenance index database," May 2016. [Online]. Available: http://www.gettyimages.com/