2021 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)

Social Crowd Simulation: The Challenge of Fragmentation

Michelangelo Diamanti
Center for Analysis and Design of Intelligent Agents
Department of Computer Science
Reykjavik University
Reykjavik, Iceland
michelangelo17@ru.is

Hannes Högni Vilhjálmsson
Center for Analysis and Design of Intelligent Agents
Department of Computer Science
Reykjavik University
Reykjavik, Iceland
hannes@ru.is

***Abstract*—Crowd simulation is the process of simulating the movement and behavior of a large number of people. This field is continuously being improved by incorporating different theories of how humans move and interact with their surroundings, steadily increasing the realism of the simulation. Furthermore, new techniques for calibrating simulation parameters, and evaluating the accuracy of simulation output, keep being proposed. This paper presents a brief overview of these foundations and argues that a fragmentation of the field into multiple incompatible solutions may impede progress towards comprehensive social behavior models. It finally argues that abstractions of human intent and behavior, proposed within the Embodied Conversational Agents community, may suggest a useful path towards bringing social crowds to new levels of realism.**

***Index Terms*—crowd simulation, social behavior, parameter calibration, model evaluation, unified framework**

I. INTRODUCTION

Crowd simulation is the process of simulating the movement and the behavior of a large number of people [1]. The majority of crowd simulators envision humans as individualistic entities, who navigate their surroundings to achieve goals while avoiding obstacles. Among the set of typical obstacles they try to avoid, are the other humans. On the other hand, the term *social* crowd simulation refers to simulations that treat other humans as something more than mere obstacles, and based on theories of human social behavior, see them as potential interaction partners. These theories provide explanations for how human behavior is influenced by the behavior of other humans, as well as by the social or psychosocial context in which the behavior takes place. Realistic crowd simulation has become a fundamental research topic in computer science due to its potential useful application in many different fields. For example, it has been employed for social service training and treating social phobias [2]. Another classic application is in urban planning, where it can help predict the flow and behavior of people in public spaces [3]–[5]. Moreover, crowd simulation has been very useful for replacing background actors in movies [6], and for populating video game environments [7]. All these applications would benefit from agents capable of social reasoning, as social solutions tend to make the crowds seem more believable [8], [9]. Thus, there is a need for platforms that make it easier to adopt social models when populating virtual worlds. This is even more urgent for immersive platforms such as Virtual/Augmented Reality, which often require higher levels of realism to maintain the suspension of disbelief. These requirements are partly satisfied by Game Engines which facilitate the integration of high level behaviors, navigation, animation, and rendering.

In addressing this need, researchers have successfully applied a range of social behavior theories, studied in psychology and sociology, for generating and animating virtual human behavior in crowd simulations. These behavior models sit on top of a variety of navigation algorithms that let the virtual humans traverse the virtual space. These simulations are also being pushed towards greater realism through various ways of objectively calibrating simulation parameters. Finally, different ways to evaluating simulation models and their outputs are constantly being invented.

The social simulation field therefore rests on a foundation that consists of a number of sub-fields, each proposing a specific approach and technical solution. This paper presents a brief overview of each of these sub-fields, and then provides a specific example of how challenging it can be to combine state-of-the-art solutions. It concludes by arguing that the fragmentation of crowd simulation research and technology may be impeding progress, and that a common framework, and a point of reference, may be useful.

II. RELATED WORK

The crowd simulation field rests on a foundation that consists of a number of sub-fields, each proposing a specific approach and technical solution: navigation algorithms, social behavior models, calibration and evaluation techniques. We will now quickly give some definitions and examples concerning them.

Navigation Algorithms. A navigation algorithm is essentially a solution for making an agent successfully traverse an environment, from a starting position to a destination position, without getting stuck. They usually employ two components, one for global path planning which returns a complete route from a starting point to a destination (e.g. navigation meshes [10]), and the other for local obstacle avoidance which deals with the dynamic part of the environment

978-1-6654-3225-2/21/$31.00 ©2021 IEEE
DOI 10.1109/AIVR52153.2021.00034

that cannot be handled by a single computation (e.g. velocity obstacles [11]). One particular approach [12] tries to minimize the undesired motion stemming from disagreements between global path planning and local collision avoidance. The authors devised a technique for generalizing path planning trajectories and velocities into strategies. Each strategy is a generalized representation which can be compared and combined with other strategies. For example, an agent may re-plan its global path when collision avoidance suggests a detour, leading to smoother trajectories.

Social Behavior Models. In social crowd simulations, humans are considered more than mere obstacles to avoid, but rather as social entities that can be interacted with or have social or psychological influence. Advances in this field have relied on theories from sociology and psychology such as: Social Group Models [13], Social Emotion Models [14], and Social Activity Models [15]. By contrast, several crowd simulations are based upon rules more akin to physics models (such as fluid dynamics), rather than social ones. These simulations, although representing crowd movements, do not account for the social aspect, and are thus not part of social crowd simulation.

Calibration Techniques. As crowd simulation tries to emulate something as complex as human beings, the model needs to capture many of their features through a range of parameters. Thus, there is a clear need for a method that finds the best value for every parameter of the simulation, for producing acceptable results. Some approaches that can solve this task are: Search Formulation [16], Optimization Approaches [17], and Evolutionary Algorithms [18].

Evaluation Techniques. With increasing focus on the realism offered by crowd simulation, there is a growing need to evaluate them against real world data. Some examples of available objective evaluation techniques include: Fundamental Diagram [19], Entropy Metric [20], Edit Distance Metric [21], Trending Paths [22]. Particularly noteworthy is [23], which proposes an approach for selecting the pedestrian model that best simulates a given scenario. The authors employ a technique similar to the entropy metric for evaluating the difference between simulated and real data. Moreover, they employ manifold learning to craft a general representation of crowd states which they call crowd space. This encoding provides a natural way to represent a wide range of crowd scenarios, where similar crowd scenarios lie near each other on the manifold. By combining these two techniques, it is possible to assess how well any given pedestrian model can simulate a certain scenario. Then, there are also subjective approaches which rely on user studies for evaluating the simulations [24]. These approaches are often preferred in highly immersive environments such as in VR/AR.

Crowd Simulation Frameworks. Given that there are several research areas contributing to the advancement of crowd simulation, and that a ”divide and conquer” approach can provide good results, it is only natural that the field is getting fragmented. This fragmentation may be related to the different scientific backgrounds and objectives of research groups, as argued in [25]. The authors explain that crowd simulation is an interdisciplinary field investigated by different scientific communities. Each community may have its own set of methods and techniques for studying pedestrian dynamics and, sometimes, they take the same term to mean different things. However, the goal is always the best possible behavior overall, hence the idea of unification. The idea of unifying crowd simulation has already been explored in the literature. Vadere [25] is one example of such effort, as it offers a common platform containing several simulation models. End users can simply pick and chose among these models to simulate pedestrian dynamics, while developers can create new models and even extend the framework. Moreover, Vadere offers a convenient GUI to visualize the simulation process, and extract relevant information for analyzing the results. Another notable example is a crowd simulation framework called Menge [26]. The framework aims at separating crowd simulation into decoupled sub-problems, whose solution can then be more easily reused by other members of the community. First, *goal selection* involves determining what each agent wants to achieve based on several factors such as psychology and world knowledge, this is achieved through Behavioral Finite State Machine (BFSM). Second, *plan computation* means devising a sequence of actions for reaching the chosen goal, by employing several techniques such as Navmesh, road maps, and velocity fields based approaches. Finally, there is *plan adaptation* which adjusts the previously computed plan to account for dynamic phenomena, based on one of several available pedestrian models. These abstractions allow researchers to focus on a single aspect of the simulation model, delegating the complexity overhead of the remaining components to the framework itself. It then becomes possible to compare simulation models at a more granular level, since any of the components can be specifically matched against its alternative. However, these frameworks have some important limitations: they do not offer any built in solutions to calibrate the model parameters, nor to evaluate simulation outputs. This makes it difficult to perform meaningful comparisons between components. Even more importantly, they do not support the combination of high level behaviors, which is a required feature towards a comprehensive model of human social behavior.

III. AN EXAMPLE OF FRAGMENTATION

While the foregoing brief overview of the social crowd simulation sub-fields of *navigation techniques*, *social modeling*, *calibration techniques* and *evaluation techniques*, has revealed creative and vibrant research activity, it has also highlighted how vastly different, and in some ways, incompatible the contributions to the field have been. This is a worrying trend of fragmentation that may threaten steady progress. As a case study, to drive this point home, we will now compare three influential, but different, papers which model group behavior. They each propose a valuable and complementary model of behavior, but bringing them together is fundamentally challenging due to incompatible technical foundations.

The first paper [27] models group behavior by relying on a leader-follower mechanic. Each group has a predefined single leader agent that is moving towards a target position. Every other agent in the group adapts its velocity vector and orientation to follow the leader: if it gets too close it decelerates, while if it's falling behind it accelerates to keep the pace. The model was implemented on top of a Cellular Automaton navigation algorithm, which discretizes space into cells and allows agents to move between them. The authors didn't specify any particular calibration nor evaluation technique.

The work of [28] proposes a model that captures both how agents in the same group influence each other (intra group dynamics), and how one group can influence the behavior of other groups (inter groups dynamics). The aforementioned dynamics are specified by two-dimensional matrices which also denotes, for every entry, the strength of the influence. The intra group matrix is used to compute, for each agent, its distance from the center of the group and the average moving direction based on other agents that have a non-zero influence. Each agent will then adjust its trajectory to maintain the group structure. In a similar way the leader of each group can be influenced by the behavior of other groups – thus indirectly influencing its followers. The model is implemented through Reynolds's Steering Behaviors with three main behaviors: (i) moving to a randomly generated destination; (ii) avoiding collision with obstacles; (iii) maintaining the intra-group structure and the inter-group relationship. The parameters of the model have been tweaked to discover their connection to different group shapes and crowd behaviors. Moreover, the model was employed to simulate a test scenario, and the results have been compared with predictions of density/flow ratio to evaluate its performance.

Finally, [29] employ a formal description of gaze direction to model social group behavior. The idea is that gaze direction and eye contact are essential features of group communication, and that pedestrians adjust their position to maintain a comfortable posture allowing them to socialize. This idea is modeled adjusting the velocity vector of each agent by a factor inversely proportional to its gaze direction with respect to its partners. The model was implemented extending Helbing social force model [30] with a new term which represents the ratio between acceleration and gaze rotation. Although the authors did not calibrate the parameters of the extended version of the model, they claim that the original one was already calibrated in previous work [29]. On the other hand the new model's accuracy was evaluated by comparison of walking patterns coming from the model and real-world data. In particular two kind of analyses were performed: the first one computed the average angle and distance between each pair of pedestrians, and studied the related organization patterns. The second one was an ANCOVA (Analysis of Covariance) test which confirmed a linear decrease of walking speeds with increasing group size – as observed in the real data.

It is possible to see that each one of the aforementioned works proposes a social behavior model which addresses a single important aspect of group behavior – essentially *one piece* of the overall puzzle: leader influence, group influence and gaze influence. This is due to the difficulty of creating a comprehensive simulation model for human behavior, which has led researchers to split the challenge into various sub-problems. This divide-and-conquer approach has allowed them to focus on a single trait of human behavior at a time, restricting the set of parameters for each experimental design, and making the challenge more manageable. At the same time, each of these works is implemented, calibrated and evaluated using different incompatible techniques. Let us take as an example the navigation algorithm used by each of the models. The three social theories modelled there all fall in the category of group behavior, yet they have different implementation details. Both [27] and [29] base their navigation implementation on the social force model [30], introducing in their systems attractive and repulsive forces, which drive the agents movements. However, the output of the navigation system of [27] is a new cell in the cellular automaton, whereas the output in [29] is a new agent position in continuous space. Thus, combining these two models would require merging two paradigms for spatial representation. Moreover, adding [28] to the mix would not be any easier, since it relies on the third paradigm of continuous velocity updates through classic Reynolds steering behaviors.

For these reason it would be rather difficult to perform a fair comparison among these models. And it would be even more difficult to merge them together to see whether a more comprehensive and realistic behavior emerges. As a result, crowd simulation research might not be advancing as quickly as it could be, if researchers were able pool their resources to build more sophisticated crowd models.

IV. PROPOSED STEPS

A relevant example of how unification can benefit research can be observed in the Embodied Conversational Agents (ECA) community. Researchers produced a unified framework [31] which splits ECA behavior generation into three independent stages: (i) intent planning, (ii) behavior planning, and (iii) behavior realization. A Function Markup Language (FML), describing intent without referring to physical behavior, mediates between the first two stages and a Behavior Markup Language (BML) describing desired physical realization, mediates between the last two stages. This allowed researchers to focus on modeling high level intent planning for ECAs, without worrying about available behaviors, and on behavior specification, without worrying about their underlying realization or animation capabilities. Among other things, researchers have employed the framework to create standard character animation engines, essentially behavior realizers, that are being shared within the ECA community [32], fostering collaboration and accelerating progress in the field. A similar approach could be employed to crowd simulation, by creating a common framework for bridging the several sub-fields that are shaping its progress. But since the field of crowd simulation is somewhat different from that of ECAs, there will be some unique challenges to tackle. For example, since crowd simulation deals with a large number of agents, special attention needs to be paid to performance. In fact, resources allocated to each agent must be limited, to ensure that the system can run smoothly. This also highlights the importance of keeping the performance overhead of a general framework to a minimum. Another peculiar challenge is tied to crowd variation. Given the need of simulating a multitude of agents, they need to be different both in aspect and in behavior, otherwise they would appear identical and odd.

The idea of a common framework has been in part realized in Menge, where high level behavior specifications are written in XML, and then the realization is delegated to the system which can employ several navigation algorithms and models. So, one idea would be to build upon Menge to make it capable of merging crowd models into a unified simulation of human social behavior. The first step towards this would be investigating to what extent Menge can support the modelling of different kinds of crowd simulations. The next goal would be to extend Menge by integrating state-of-the-art calibration and evaluation techniques inside the framework. This would be accomplished by creating interfaces that allow fetching information about the simulation state where necessary. Then, it would be possible to create a mechanism which allows to blend different theories of human behaviors into a unified model. One possible representation of which would allow this blending are heat-maps, each one holding information about a particular aspect of human behavior, in relation to the the environment. The contents of these data structures can then be blended and queried by the pedestrian models, effectively resulting in a behavior dictated by several theories of human behavior. In some instances, it might not be possible to blend the behaviors, for example when they're too different from each others. A solution in these cases, would be to implement arbitration techniques which pick the behavior most appropriate for the context in play. Additionally, the previously mentioned extensions to the framework will make it possible to compare the simulation output with reference evaluation datasets, to objectively assess the realism of the models. Thus, as the simulation is carried out, it will be possible to run calibration algorithms and chose the parameters which draw the simulation closest to reference data.

V. CONCLUSION

This paper has argued that the field of social crowd simulations sits on four different sub-fields that represent different dimensions of growth: *navigation algorithms*, *social behavior theories*, *calibration techniques*, and *evaluation techniques*. Research in these sub-fields has produced a range of useful techniques that can be employed for improving crowd simulation. At the same time, this abundance of different approaches has led to fragmentation in crowd simulation research overall. When crowd simulation researchers experiment with social behavior theories for replicating human movements, they usually draw from sociology and psychology studies. Then, they extend one of the available navigation algorithms and create a simulation model. Finally, they may arbitrarily chose a calibration and evaluation technique to reduce the gap between the simulation and real-world data. Since these components vary greatly from one research effort to another, it is challenging to objectively compare different models – and it is even more difficult to combine multiple models into a stronger overall comprehensive model of human behavior.

It is possible that crowd simulation research would advance more rapidly if experts were more easily able to share and build on top of each other's modules and findings. Menge is one example of a framework which started addressing this problem, but there are still several important challenges which need to be tackled in order to support the modeling of comprehensive human social behavior: (i) it does not offer any built in solutions to calibrate the model parameters, nor to evaluate simulation outputs, and (ii) it does not support the combination of high level behaviors. By leveraging off existing advances within the Embodied Conversational Agents community, in particular efforts to create useful abstractions of human behavior, we may yet be able to bring social crowds to new levels of realism.

VI. ACKNOWLEDGEMENTS

This work has been supported by two grants from the Icelandic Research Fund (#217663-051, #174444-053). The authors would like to thank Claudio Pedica for thoughtful discussions and valuable insights.

REFERENCES

[1] D. Thalmann, "Crowd simulation," *Wiley Encyclopedia of Computer Science and Engineering*, 2007.

[2] D. Opriş, S. Pintea, A. García-Palacios, C. Botella, Szamosközi, and D. David, "Virtual reality exposure therapy in anxiety disorders: a quantitative meta-analysis: Virtual Reality Exposure Therapy," *Depression and Anxiety*, vol. 29, pp. 85–93, Feb. 2012.

[3] C. D. T. Mathew, P. R. Knob, S. R. Musse, and D. G. Aliaga, "Urban Walkability Design Using Virtual Population Simulation," *Computer Graphics Forum*, vol. 38, pp. 455–469, Feb. 2019.

[4] T. Feng, L.-F. Yu, S.-K. Yeung, K. Yin, and K. Zhou, "Crowd-driven mid-scale layout design," *ACM Transactions on Graphics*, vol. 35, pp. 1–14, July 2016.

[5] O. Rogla, N. Pelechano, and G. A. Patow, "Procedural Crowd Generation for Semantically Augmented Virtual Cities," *Computers & Graphics*, vol. 99, pp. 83–99, 2021.

[6] R. Scott, "The Lord of The Rings: The Two Towers," *SIGGRAPH*, p. 5, 2003.

[7] T. F. Cournoyer and A. Fortier, "Massive crowd on ACU: AI Recycling."

[8] C. Peters and C. Ennis, "Modeling Groups of Plausible Virtual Pedestrians," *IEEE Computer Graphics and Applications*, vol. 29, pp. 54–63, July 2009. Conference Name: IEEE Computer Graphics and Applications.

[9] C. Pedica, H. H. Vilhjálmsson, and M. Lárusdóttir, "Avatars in conversation: the importance of simulating territorial behavior," in *International Conference on Intelligent Virtual Agents*, pp. 336–342, Springer, 2010.

[10] X. Cui and H. Shi, "An Overview of Pathfinding in Navigation Mesh," *IJCSNS*, vol. 12, no. 12, p. 4, 2012.

[11] J. van den Berg, M. Lin, and D. Manocha, "Reciprocal Velocity Obstacles for real-time multi-agent navigation," in *2008 IEEE International Conference on Robotics and Automation*, pp. 1928–1935, May 2008. ISSN: 1050-4729.

[12] W. v. Toll and J. Pettré, "Synchronizing navigation algorithms for crowd simulation via topological strategies," *Computers & Graphics*, vol. 89, pp. 24–37, 2020.

[13] C.-T. Li and S.-D. Lin, "Social flocks: a crowd simulation framework for social network generation, community detection, and collective behavior modeling," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '11, (San Diego, California, USA), pp. 765–768, Association for Computing Machinery, Aug. 2011.

[14] Y. Mao, S. Yang, Z. Li, and Y. Li, "Personality trait and group emotion contagion based crowd simulation for emergency evacuation," *Multimedia Tools and Applications*, May 2018.

[15] W. Shao and D. Terzopoulos, "Autonomous pedestrians," *Graphical Models*, vol. 69, pp. 246–274, Sept. 2007.

[16] F. Azadivar, "SIMULATION OPTIMIZATION METHODOLOGIES," *Proceedings of the 31st conference on Winter simulation: Simulation a bridge to the future-Volume 1*, p. 8, 1999.

[17] D. Wolinski, S. J. Guy, A.-H. Olivier, M. Lin, D. Manocha, and J. Pettré, "Parameter estimation and comparative evaluation of crowd simulations," *Computer Graphics Forum*, vol. 33, no. 2, pp. 303–312, 2014.

[18] Z. Jin and B. Bhanu, "Optimizing crowd simulation based on real video data," in *2013 IEEE International Conference on Image Processing*, pp. 3186–3190, Sept. 2013. ISSN: 2381-8549.

[19] A. Seyfried, B. Steffen, W. Klingsch, and M. Boltes, "The Fundamental Diagram of Pedestrian Movement Revisited," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2005, pp. P10002–P10002, Oct. 2005. arXiv: physics/0506170.

[20] S. J. Guy, J. van den Berg, W. Liu, R. Lau, M. C. Lin, and D. Manocha, "A statistical similarity measure for aggregate crowd dynamics," *ACM Transactions on Graphics*, vol. 31, p. 1, Nov. 2012.

[21] L. Chen, M. T. Özsu, and V. Oria, "Robust and fast similarity search for moving object trajectories," in *Proceedings of the 2005 ACM SIGMOD international conference on Management of data*, SIGMOD '05, (Baltimore, Maryland), pp. 491–502, Association for Computing Machinery, June 2005.

[22] H. Wang, J. Ondřej, and C. O'Sullivan, "Trending Paths: A New Semantic-Level Metric for Comparing Simulated and Real Crowd Data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, pp. 1454–1464, May 2017.

[23] I. Karamouzas, N. Sohre, R. Hu, and S. J. Guy, "Crowd space: a predictive crowd analysis technique," *ACM Transactions on Graphics*, vol. 37, pp. 1–14, Jan. 2019.

[24] F. A. Rojas and H. S. Yang, "Immersive human-in-the-loop HMD evaluation of dynamic group behavior in a pedestrian crowd simulation that uses group agent-based steering," in *Proceedings of the 12th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry - VRCAI '13*, (Hong Kong, Hong Kong), pp. 31–40, ACM Press, 2013.

[25] B. Kleinmeier, "Vadere: An Open-Source Simulation Framework to Promote Interdisciplinary Understanding," *Collective Dynamics*, vol. 4, pp. 1–34, 2019.

[26] S. Curtis, A. Best, and D. Manocha, "Menge: A Modular Framework for Simulating Crowd Movement," *Collective Dynamics*, vol. 1, p. A1, Mar. 2016.

[27] M. Seitz, G. Köster, and A. Pfaffinger, "Pedestrian Group Behavior in a Cellular Automaton," in *Pedestrian and Evacuation Dynamics 2012* (U. Weidmann, U. Kirsch, and M. Schreckenberg, eds.), pp. 807–814, Cham: Springer International Publishing, 2014.

[28] F. Qiu and X. Hu, "Modeling group structures in pedestrian crowd simulation," *Simulation Modelling Practice and Theory*, vol. 18, pp. 190–205, Feb. 2010.

[29] M. Moussaïd, N. Perozo, S. Garnier, D. Helbing, and G. Theraulaz, "The Walking Behaviour of Pedestrian Social Groups and Its Impact on Crowd Dynamics," *PLoS ONE*, vol. 5, p. e10047, Apr. 2010.

[30] D. Helbing and P. Molnár, "Social force model for pedestrian dynamics," *Physical Review E*, vol. 51, pp. 4282–4286, May 1995.

[31] S. Kopp, B. Krenn, S. Marsella, A. N. Marshall, C. Pelachaud, H. Pirker, K. R. Thórisson, and H. H. Vilhjálmsson, "Towards a Common Framework for Multimodal Generation: The Behavior Markup Language," in *Intelligent Virtual Agents: 6th International Conference, Marina Del Rey, CA, USA, August 21-23, 2006. Proceedings* (J. Gratch, R. M. Young, R. Aylett, D. Ballin, and P. Olivier, eds.), IVA '06, pp. 205–217, Berlin, Heidelberg: Springer, 2006.

[32] H. Vilhjálmsson, N. Cantelmo, J. Cassell, N. E. Chafai, M. Kipp, S. Kopp, M. Mancini, S. Marsella, A. N. Marshall, C. Pelachaud, Z. Ruttkay, K. R. Thórisson, H. van Welbergen, and R. J. van der Werf, "The Behavior Markup Language: Recent Developments and Challenges," in *Intelligent Virtual Agents* (C. Pelachaud, J.-C. Martin, E. André, G. Chollet, K. Karpouzis, and D. Pelé, eds.), vol. 4722, pp. 99–111, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007. Series Title: Lecture Notes in Computer Science.