Behavioral Landmarks: Inferring Interactions from Data

M. Lemonari[1], P. Charalambous[2], A. Panayiotou[1,2], Y. Chrysanthou[1,2], J. Pettré[3]

[1]University of Cyprus, Cyprus, [2]CYENS - Centre Of Excellence, Cyprus, [3]Inria, University of Rennes, CNRS, IRISA, France

Abstract

We aim to unravel complex agent-environment interactions from trajectories, by explaining agent paths as combinations of pre-defined basic behaviors. We detect trajectory points signifying environment-driven behavior changes, ultimately disentangling interactions in space and time; our framework can be used for environment synthesis and authoring, shown by our case studies.

CCS Concepts

• ***Computing methodologies*** → ***Image representations; Neural networks;*** *Motion processing;*

1. Introduction and Related Work

In crowd simulation the focal point is to understand how people behave in public i.e., interacting with each other and the environment. In line with this, our objective is to disentangle the way agents interact with different points in a given environment, by observing their trajectories. Ultimately, having knowledge of how people move in an environment, would enable creators to replicate [JCP*], and eventually create new agent simulations, with different environment layouts, while maintaining their interactions.

Crowd simulation has been tackled from several perspectives [VTP21]. Specific to our objective, methods attempt to extract optimal parameters out of given agent trajectories [WJGO*14]. However, these are generally limited to navigation parameters [KSH*, KSHG18]. Other works try to capture relations between motion and environment, or people [ST05]. Also, distributing the simulation parameters spatially and temporally, is restricted to pre-defined grid divisions and time windows, usually chosen based on trial-and-error or simple rule-books; observing real behaviors reveals this to be sub-optimal. Our aim is towards simulations capable of richer behavior descriptions, where complex interactions with the environment are captured e.g., an agent trajectory in a train station, is not reduced to just going from the entrance to the train, but rather, also passing through shops, or interacting with vending machines. Such behaviors result in specific trajectories, reflecting an underlying intricate pattern of environment interactions. Our work differs from SOTA, in that we try to infer complex relations from trajectories only, without object knowledge.

Attempting to understand how agents interact with environmental clues from their trajectories, we propose a learning-based method that leverages the concept of Interaction Fields (InFs) [CvTH*21]. Colas et al. demonstrate a simple way to capture basic agent/environment interactions such as approaching, or circling around, each of them corresponding to a velocity field.

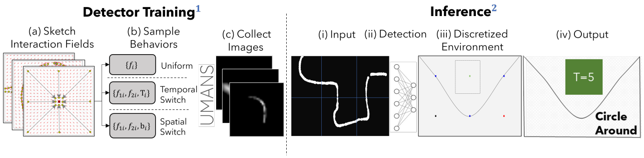

Figure 1: *Training and Inference Pipeline Overview.*

Taking advantage of the intuitiveness and usability of this work, we view trajectories as a combination of InFs in order to unravel the behavior dynamics both in space and time. Drawing inspiration from recent works, utilizing Machine Learning (ML) to infer crowd characteristics [CPV*23], we design a neural network to detect where and when simulation parameters change, thus helping artists author more informed crowds. In a nutshell, we harness the ML potential to understand environment-guided agent movements, delivering a more robust environment discretization, based on InFs.

2. Terminology and Pipeline

Terminology: In the pursuit of inferring agent/environment interactions based on trajectories, we define specific trajectory points, "Landmarks" (LMs), to act as indicators of environment-affected behavior changes e.g., from *avoiding* to *approaching*. We distinguish between: **Spatial** LMs, reflecting agents changing behavior due to the area entered e.g., walking from the street into an exhibition room, and **Temporal** LMs, reflecting behavior change due to a timed switch/event e.g., a celebrity walked into the room. We speculate that temporal LMs would correspond to abrupt path changes, whereas spatio-temporal clues would reveal a spatial LM e.g., smooth transitions and lower speeds. Thus, we presume that such differences are detectable in trajectory images, and hence we design an image-based learning strategy. **Pipeline Overview:** We detect the existence and type of LMs by training a Convolutional Neural Network (CNN). For training, we generate trajectory im-

DOI: 10.2312/egp.20241039

ages based on combinations of four basic InFs: "Approach", "Wander", "Circle Around" and "Avoid", signifying either temporal LM, spatial LM, or uniform basic behavior (no LM presented). After training the detector on synthetic data, we develop a sketching interface where users can arbitrarily draw crowd paths. Applying our model on the sketched paths, gives the discretized environment, facilitating new layouts and distributing user-collected landmark prefabs to populate them.

Detector Training Data: We generate ~64K synthetic images, each corresponding to either uniform behaviors, spatial, or temporal LMs. For generating LM images, we randomly sample two distinct basic behaviors from the selected InFs. For temporal and spatial LMs, we select a random temporal switch $T \in \{1,\ldots,9\}$ and blending weight $b \in \{1,\ldots,5\}$, respectively. We simulate using UMANS [vTGG*20] and use the exported trajectories to create images, encoding both agent path and speed; higher pixel intensity corresponds to lower speed values (Figure 2).

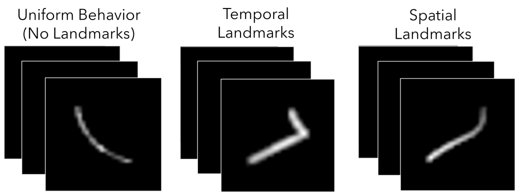

Figure 2: *Generated training images along with their class type.*

Models: We train a *LM Detector* (3-class image-based CNN) to detect the existence and type of possible LMs (Figure 1-left). We test on synthetic data, yielding an accuracy of 84.67%. We also train a set of *Auxiliary Models*, an identifier and a temporal classifier extending the capabilities of our system. Given the case of uniform behaviors, the *LM identifier* finds the behavior (InF) from the predefined set of InFs. Given the presence of a temporal LM, the *classifier* finds the time of the switch; code available here.

Inference: Users can arbitrarily sketch crowd flow(s) in an area that they define (Figure 1-right). The drawing speed defines the path line's thickness to comply with the training data (lower speed corresponds to bigger points). The detector then finds the landmark points, based on which we heuristically discretize the space.

3. Results

We conduct scenario-specific case studies to assess our framework's plausibility by inputting user sketches and authoring novel environments and movements. We ask an experienced and an inexperienced user to draw 5 sketches. Firstly, we give them a scenario ("park", "museum", "street" or "train station"), and ask them to sketch the environment and corresponding crowd flow they have in mind. We then collect their description of their intended crowd behavior e.g., "People walking around in a park while others are stationary around a kiosk". Figure 3 shows example sketches along with the discretized environment (middle), which can be used to **synthesize** a new scene, informed by the behavior areas implicitly defined by the user, via our LM detection. Then, they are asked to design the crowd flow of *timed events* e.g., "at some point a celebrity walks in", and *area constraints* e.g., "there is one room with exhibits and another showing a documentary", thus collecting temporal and spatial LM prefabs (100% & 80% prediction accuracies). We distribute the prefabs guided by the synthesized environment, hence **authoring** new simulation with different flows (Figure 4).

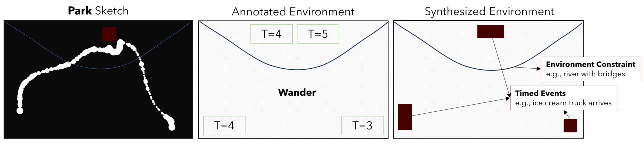

Figure 3: *Landmark-based area discretization on user sketch.*

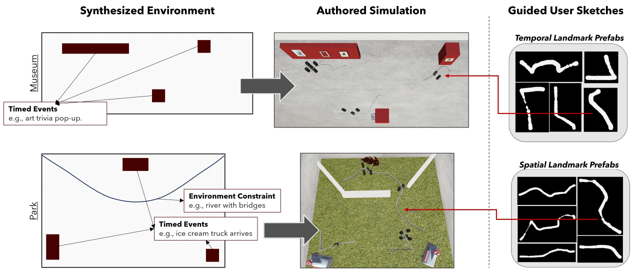

Figure 4: *Authoring using discretized area and landmark prefabs.*

4. Future Work

We aim to improve the generalisability of our framework by training on multiple-sized images and evaluate on real data, even though it is difficult to obtain ground truths. We are also interested in extending this to investigate social agent interactions, and specified environment-agent interactions e.g., ticket machines.

Acknowledgements

This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska Curie grant agreement No 860768 (CLIPE project).

References

[CPV*23] CHARALAMBOUS P., PETTRE J., VASSILIADES V., CHRYSANTHOU Y., PELECHANO N.: Greil-crowds: Crowd simulation with deep reinforcement learning and examples. *ACM TOG* (2023). 1

[CvTH*21] COLAS A., VAN TOLL W., HOYET L., PACCHIEROTTI C., CHRISTIE M., ZIBREK K., OLIVIER A.-H., PETTRÉ J.: Interaction fields: Sketching collective behaviours. In *CGF* (2021). 1

[JCP*] JU E., CHOI M. G., PARK M., LEE J., LEE K. H., TAKAHASHI S.: Morphable crowds. *ACM TOG*. 1

[KSH*] KAPADIA M., SINGH S., HEWLETT W., REINMAN G., FALOUTSOS P.: Parallelized egocentric fields for autonomous navigation. *The Visual Computer*. 1

[KSHG18] KARAMOUZAS I., SOHRE N., HU R., GUY S. J.: Crowd space: a predictive crowd analysis technique. *ACM TOG* (2018). 1

[ST05] SHAO W., TERZOPOULOS D.: Autonomous pedestrians. In *Proceedings of the 2005 ACM SIGGRAPH/SCA* (2005). 1

[vTGG*20] VAN TOLL W., GRZESKOWIAK F., GANDÍA A. L., AMIRIAN J., BERTON F., BRUNEAU J., DANIEL B. C., JOVANE A., PETTRÉ J.: Generalized microscropic crowd simulation using costs in velocity space. In *I3D* (2020). 2

[VTP21] VAN TOLL W., PETTRÉ J.: Algorithms for microscopic crowd simulation: Advancements in the 2010s. In *CGF* (2021). 1

[WJGO*14] WOLINSKI D., J. GUY S., OLIVIER A.-H., LIN M., MANOCHA D., PETTRÉ J.: Parameter estimation and comparative evaluation of crowd simulations. In *CGF* (2014), Wiley Online Library. 1