Data Driven Evaluation of Crowds

Alon Lerner[1], Yiorgos Chrysanthou[2], Ariel Shamir[3], and Daniel Cohen-Or[1]

[1] Tel Aviv University, Israel
[2] University of Cyprus, Cyprus
[3] The Interdisciplinary Center, Israel

Abstract. There are various techniques for simulating crowds, however, in most cases the quality of the simulation is measured by examining its “look-and-feel”. Even if the aggregate movement of the crowd looks natural from afar, the behaviors of individuals might look odd when examined more closely. In this paper we present a data-driven approach for evaluating the behaviors of individuals within a simulated crowd. Each decision of an individual agent is expressed as a state-action pair, which stores a representation of the characteristics being evaluated and the factors that influence it. Based on video footage of a real crowd, a database of state-action examples is generated. Using a similarity measure, the queries are matched with the database of examples. The degree of similarity can be interpreted as the level of “naturalness” of the behavior. Essentially, this sort of evaluation offers an objective answer to the question of how similar are the simulated behaviors compared to real ones. By changing the input video we can change the context of evaluation.

1 Introduction

The simulation of computer generated crowds has progressed in leaps and bounds from its conception more than two decades ago. In films, computer games and other virtual world applications we have seen simulations of every type of crowd imaginable, from flocks of dinosaurs to plain everyday pedestrians. Despite the fact that these crowds differ in size, behavior and method of simulation, the common goal remains the same; to simulate the most naturally looking crowd possible.

Different types of crowds may present different types of natural behaviors. Pedestrians do not behave nor look like people playing sports. How would one know if the simulation looks natural? The only means for evaluating the results of a simulation so far were subjective, e.g. looking at it and deciding whether it “looks natural”. In general, this is a valid approach for the global “look-and-feel” of the simulation.

A crowd is defined by a large number of simulated individuals. In order to assess a simulated crowd’s quality in full, it must be examined in detail, since a single individual whose behavior deviates from some “regular” or “normal” pattern may hinder the quality of the entire simulation. Even a human paying close attention to a simulation will find it difficult to measure the quality of all individual behaviors at all times. And so the need for an effective automatic measure becomes a necessity. Evaluating individual behaviors also depends on one’s interpretation of the basic notion of how people behave. Moreover, multiple interpretations are valid, and different people may choose to focus on different aspects of the same behavior. How can we define what is “normal” or “natural”?

A. Egges, R. Geraerts, and M. Overmars (Eds.): MIG 2009, LNCS 5884, pp. 75–83, 2009.

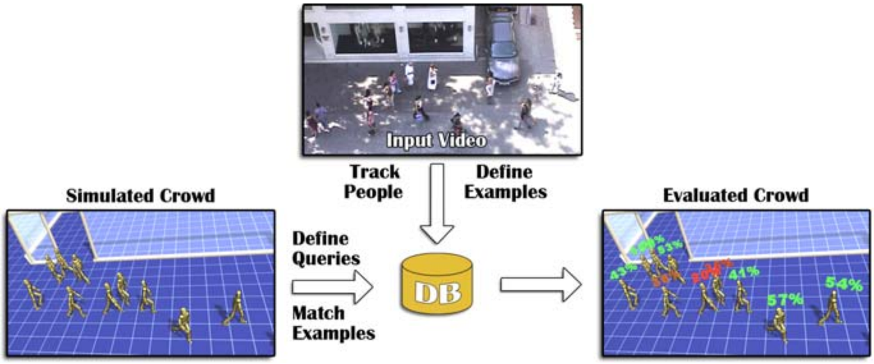

Fig. 1. Overview of context dependant evaluation: a specific context is defined by an input video. The video is analyzed and a database of examples is created. For a given simulation, each individual trajectory is analyzed and compared to the "normal" behaviors defined by the examples in the database. The results provide an evaluation score from 0 to 100 for each individual at any time (the numbers above each individual).

In this paper we utilize a data-driven approach to assess individual behaviors within crowds. The example set, which represents individual behaviors that are considered "normal", defines the context for the evaluation. The example set is created by examining videos of real crowds and analyzing the individual trajectories and the attributes that describe the conditions under which they occurred. During an evaluation, a similar analysis is applied to the simulated trajectories, which results in a set of queries that are compared to the example set. The similarity between a query and the examples provides the grading for the simulated individual's behavior in a specific point in space and time. Such an evaluation validates "natural" behaviors and points out potentially "curious" ones within a context (see Figure 1). For instance, different types of crowds can be evaluated using different videos of crowds of a similar nature. By evaluating the behavior of *individuals* within the crowd, specific moments in time and space can be automatically identified as problematic and possibly corrected. Finally, considering the comparisons as a whole, yields a global evaluation, which reflects on the overall quality of the simulation.

Our main contribution is a data driven approach for evaluating simulated crowds based on videos of real crowds. We present results on evaluating individual behaviors within crowds for short-term intervals, although long term ones could also be evaluated using a different measure. We use different videos and different types of simulations and show examples of how "curious" behaviors could be found automatically.

2 Previous Work

Simulating crowds is a challenge that has been examined in such diverse fields as computer graphics, civil engineering, sociology and robotics. The literature is rich with a variety of different approaches, all trying to simulate a natural looking crowd, [11, 18].

Considering the diversity of the methods, particularly the ones for simulating pedestrian crowds, emphasizes the complexity of the question "what makes a crowd seem natural?" For example, some works try to minimize the amount of effort required to reach a goal, [7, 17], which originates from the notion that people are individuals that have a goal and will try to follow an "optimal" path to reach it while maintaining their speed. Other works, such as [2, 5, 19], try to simulate the flow of the crowd and pay less attention to the correctness of individual trajectories. Different works focus on different aspects of crowds. All claim to simulate natural looking crowds, and to some degree they are all correct. It is unclear which attributes are more important when behavior is concerned and since the common evaluation method is subjective, then all opinions are valid.

Although some earlier works did try to assess the quality of individual trajectories, such as [15], only recently have people begun to pay attention to the evaluation of complex multi-character environments. Ennis et al. [3] evaluate the plausibility of pedestrian formations. They manually reconstruct crowd formations from annotated photographs, modify the positions and orientations of the people and perform a perceptual evaluation of the modified static images through user experiments. It would be difficult to extend their results to animated crowds. Singh et al. [16] gain insight into the quality of a simulator using predefined test scenarios and a measure that assigns a numerical value to the average number of collisions and the time and effort required for the agents to reach their goal. Pelechano et al. [12] explore the egocentric features that a crowd simulator should have in order to achieve high levels of presence and thus be used as a framework for validating simulated crowd behaviors. Paris et al. [10] use motion capture data of interactions between pedestrians for two ends. First, they analyze the data and learn parameters for more accurate collision avoiding maneuvers. Second, they validate the results by visually comparing simulated results to the captured ones.

Data-driven techniques have been used to simulate crowds. In the work of Lee et al. [8], *state-action* examples are obtained from crowd videos. The state of an agent includes the position and movement of nearby agents, its own motion, and certain environmental features. The action is a two-dimensional vector corresponding to the agent's resulting speed and direction. In the work of Lerner et al. [9] the state stores a similar set of attributes, only at greater detail. The action is a trajectory segment that can be concatenated to a simulated agent's trajectory.

In the vision community there are many works for clustering and classifying trajectories, some also detect abnormal ones [1, 4, 6, 13, 14]. There are several significant conceptual differences between these works and ours. First, they usually do not transfer the information learned about the trajectories from one scenario to another. Second, these works, for the most part, focus on the global characteristics of the trajectories, rather than the details. Therefore, they usually classify them as a whole, rather then evaluate the quality of trajectory segments. Last, the trajectories are usually considered on their own without accounting for the stimuli that may have influenced them.

3 Overview

The underlying concept of most crowd simulators can be described by the *state-action* paradigm. For each simulated individual (*subject person*) in a specific point in time and space, a *state S* is defined as a set of potentially influential attributes, such as the person's position, speed and the direction and position of nearby individuals. Some function is then used to compare this state to predefined states, either explicitly or implicitly (e.g. a set of rules), and an *action A* is chosen and assigned to the subject person. The action can be, for example, a velocity vector or a trajectory segment.

An evaluator can be described in similar terms, aside from two key differences. First, in an evaluation process all of the state attributes, such as the full trajectories of the individuals, are known. This allows for a more accurate definition of the "natural" action that should be performed. Second, an evaluation requires a comparison between the action that should have been performed and the one that actually was performed, thus determining its quality.

An evaluation measure requires the definition of the state attributes S, the action representation A, and a similarity function between state-action pairs. The state S should include all of the attributes that potentially affect a person's decision to choose an action. For example, in the short-term measure we propose, we defined the state as the densities of the people in different regions surrounding the subject person. The action A is defined as the subject person's positions along a two second long trajectory interval.

In a preprocessing stage, the individuals observed in one or more input videos are tracked, their trajectories are analyzed and examples of state-action pairs are created. During an evaluation, the simulated trajectories are analyzed in a similar manner and *query* state-action pairs are defined. For each query, we search for the most similar examples in the example set using a similarity function. This is a two stage process. First, a set of examples whose states best match the query state are found. Then, their actions are compared. Our distance measure uses a normalized combination of both the state and action distances, in order to account for the cases where close matches for the state cannot be found. This two-stage process assures a context dependent evaluation of individual decisions, where the similarity to the closest example defines the quality of the decision embodied by the query (Figure 1).

4 Example Set and Simulations

To define a specific context for an evaluation we require a video of a crowd of a similar nature. The video should be shot from an elevated position and have a clear view of the crowd below. We implemented a simple manual tracker which requires only a few clicks of the mouse in order to track a person's trajectory. Different types of crowds are represented by different input videos (see Figure 2).

An example is defined from each time step along every trajectory of the input. We exclude only the trajectory parts in which the state and action cannot be fully defined. When adding the examples to a database, we use a greedy algorithm to cluster the examples and remove redundant ones.

We shot several videos of different crowds, Figure 2, according to which we built two databases. The first was created from two videos of a sparse crowd, whose combined length is 12 minutes and contains 343 trajectories. The second was created from a 3.5 minute long video of a dense crowd, which contains 434 trajectories. In our unoptimized implementation the time required to preprocess a sparse database was under an hour, while for a dense database over an hour. An average of about 1KB of data was stored per example.

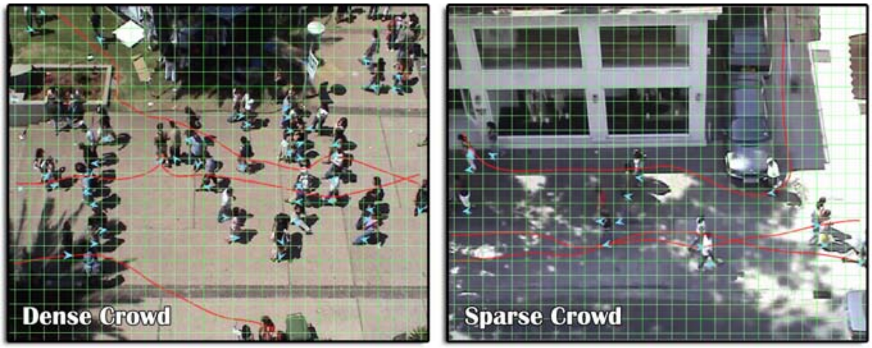

Fig. 2. Sample frames from two input videos used to define example databases. One for dense crowds and one for sparse ones. The trajectories of people (some are shown in red) were tracked manually before analysis.

To illustrate the versatility of our evaluation, we implemented two types of simulations. A simple *rule-based* simulation, where the agents maintain their speed and direction and perform only gradual turns trying to minimize collisions and an *example-based* simulation [9], which consists of agents walking on their own or in small groups. In this simulation the agents usually walk along relatively smooth trajectories, however, abrupt changes in speed or direction do appear.

5 Evaluation of Short Term Decisions

A person constantly makes short-term decisions. These are the decisions that cause him to stop, turn, slow down or speed up. These decisions are not influenced by some global objective, but rather by the local conditions surrounding the person. The impact of these decisions is immediate and short lived. Their effects can be found in short segments of a person's trajectory.

We define a measure, which we term the *density measure*, whose state attributes consist of samples of the local densities of the people surrounding the subject person. The action is defined as a two second long trajectory segment. One second before and one second after the current position. The segment is aligned such that the position and orientation of the subject person in the middle of the segment is aligned with the origin of a global coordinate system.

To define the state we divide the area surrounding the subject person into regions, Figure 3, and for each region store samples of the number of people that appear in it over a short time. This yields a compact representation of the local changes in densities in the vicinity of the subject person. The motivation for using this state definition stems from the common belief that people's reactions are influenced by the density of the people in their immediate vicinity. We define a similarity function between a query state-action pair and an example pair. The function measures the similarity between the actions (differences in positions along the trajectories) and the distance between the states (differences in densities for the surrounding regions).

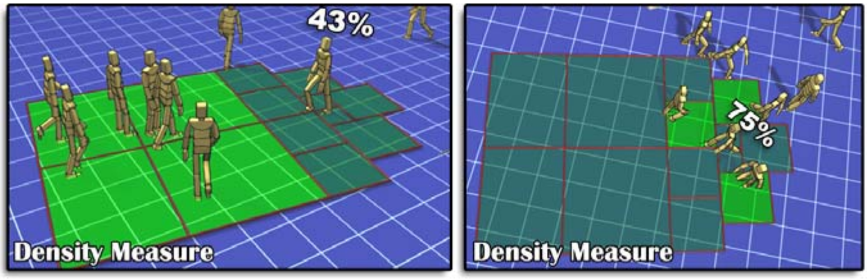

Fig. 3. In the density measure the state is composed of samples of the number of people in each one of the regions surrounding the subject person. Five samples in different times are used.

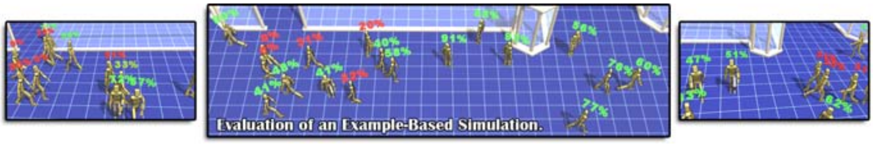

Fig. 4. A few examples from the evaluation of a crowd simulation using a sparse crowd video input and the density measure. The short-term decisions that did not find an appropriate match in the database are marked in red.

6 Results

In one of our experiments we used the database created by videos of a sparse crowd as input and evaluated the different crowds using the density measure. Results of evaluating an example-based simulation appear in Figure 4. The percentages represent the quality of the matches that were found. Zero percent means that no match was found and a hundred percent means that a perfect match was found. Low quality matches are highlighted in red. A close inspection of the evaluation results shows that, for the most part, low quality matches correspond to "curious" behaviors, traffic congestion, collisions or near misses. In Figure 4 center, the agent that received a 20% similarity value stopped walking abruptly and the four others marked in red, either walked towards an imminent collision or performed "conspicuous" evasive manoeuvres. The same can be said for most of the highlighted agents in Figure 4 left and right.

Figure 5 shows a quantitative comparison between the evaluation of the example-based simulation, the rule-based simulation and a real sparse crowd different from the input data. The columns represent ranges of similarity values (or evaluation scores), and the height of each column the relative number of queries which received a value in the range.

As can be seen, the real data received significantly better scores than the simulation. However, the example-based simulation which is collision free, contains interactions between the individuals and seems similar in nature to the input video, found a greater number of high quality matches compared to the rule-based simulation.

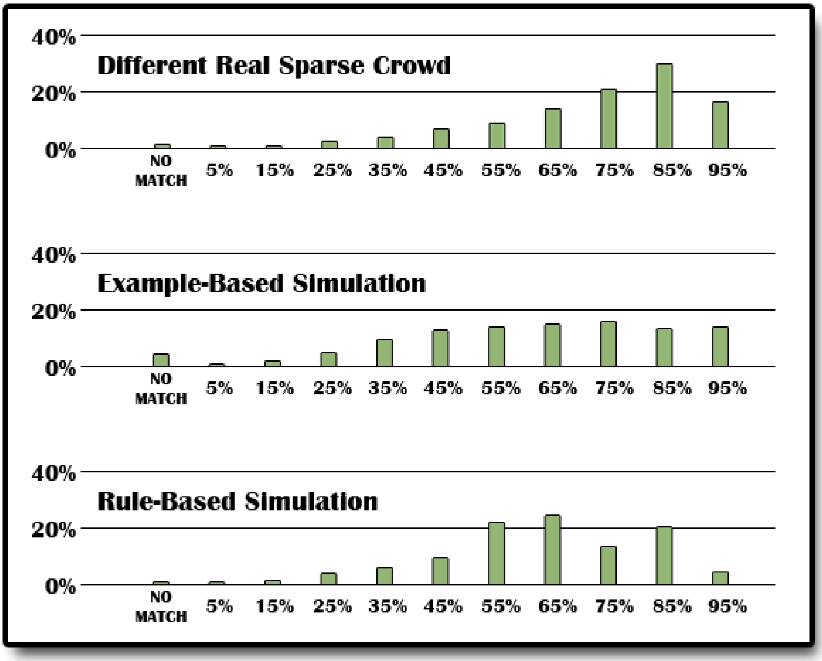

Fig. 5. A quantitative comparison between the distribution of evaluation scores when evaluating the example-based simulation, rule-based simulation and real data against the sparse crowd database. The horizontal axis represents the match values and the vertical axis the relative number of matched queries.

The most noticeable "curious" short-term decisions either involve collisions or evasive maneuvers whose goal is to avoid collisions. The method finds these and other "curious" behaviors, since similar examples were not found. On the other hand, in some cases the agents involved in collisions or near misses were deemed "natural". The reason is that a similar example where one person walked very close to another person was found in the database. Finding or not finding matches does not unequivocally mean that the behaviors are "natural" or "curious", but rather it raises the probability of them being so.

The time required to evaluate a crowd varies depending on the size of the database and the size of the simulated crowd. The simulations presented here were several minutes in length and contained a few dozen people in them. In our un-optimized implementation evaluating their short-term decisions required between 5 and 20 minutes. Using dedicated data structures to optimize the search for the best matching examples can significantly reduce the processing time.

7 Conclusions

We have presented a data-driven approach for evaluating individual behaviors in a crowd simulation. Such an approach gives the evaluation a context and a goal, since it changes the fundamental question from “how natural is this crowd?” to the relatively simpler one of “how does it compare to this real crowd?”. Using our method as a post processing stage for a simulation, it can focus the viewer’s attention on the potentially “curious” behaviors in time and space, hence reducing the effort needed for evaluation. Another significant advantage of the data-driven model is the simplicity by which the model can be changed. Changing the type of crowd which appears in the input video effectively changes the behavior model without the need to modify the evaluation application. Finally, this technique is not limited to evaluating crowds generated by a specific simulator, but rather evaluates the trajectories regardless of their source, therefore it can also be used to locate potentially “curious” behaviors in real crowds.

One disadvantage of our model is the need to obtain the right video and to track individual trajectories in it. To the best of our knowledge, to date there are no adequate automatic trackers that can accurately track people in a crowd. We used manual tracking, which is a tedious job. However, each video requires processing only once and can be used any number of times. Another limitation of the data-driven approach is the fact that if a matching example was not found in the example set it does not necessarily indicate that the behavior is “curious”. It could indeed be a natural behavior which was simply not observed in the input video.

References

[1] Brand, M., Kettnaker, V.: Discovery and segmentation of activities in video. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(8), 844–851 (2000)

[2] Chenney, S.: Flow tiles. In: Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation, pp. 233–242. Eurographics Association Aire-la-Ville, Switzerland (2004)

[3] Ennis, C., Peters, C., O’Sullivan, C.: Perceptual evaluation of position and orientation context rules for pedestrian formations. In: Proceedings of the 5th symposium on Applied perception in graphics and visualization, USA, pp. 75–82. ACM New York, NY (2008)

[4] Hu, W., Xiao, X., Fu, Z., Xie, D., Tan, T., Maybank, S.: A system for learning statistical motion patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(9), 1450–1464 (2006)

[5] Hughes, R.L.: The Flow of Human Crowds. Annual Review of Fluid Mechanics 35, 169–182 (2003)

[6] Johnson, N., Hogg, D.: Learning the distribution of object trajectories for event recognition. In: BMVC 1995: Proceedings of the 6th British conference on Machine vision, Surrey, UK, vol. 2, pp. 583–592. BMVA Press (1995)

[7] Lamarche, F., Donikian, S.: Crowd of virtual humans: a new approach for real time navigation in complex and structured environments. Comput. Graph. Forum 23(3), 509–518 (2004)

[8] Lee, K.H., Choi, M.G., Hong, Q., Lee, J.: Group behavior from video: a data-driven approach to crowd simulation. In: Proceedings of the 2007 ACM SIGGRAPH/Eurographics symposium on Computer animation, pp. 109–118 (2007)

[9] Lerner, A., Chrysanthou, Y., Lischinski, D.: Crowds by Example. Computer Graphics Forum 26(3), 655–664 (2007)

[10] Paris, S., Pettre, J., Donikian, S.: Pedestrian Reactive Navigation for Crowd Simulation: a Predictive Approach. Computer Graphics Forum 26(3), 665–674 (2007)

[11] Pelechano, N., Allbeck, J., Badler, N.: Virtual Crowds: Methods, Simulation, and Control. Synthesis Lectures on Computer Graphics and Animation. Morgan & Claypool Publishers, San Francisco (2008)

[12] Pelechano, N., Stocker, C., Allbeck, J., Badler, N.: Being a part of the crowd: towards validating VR crowds using presence. In: Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems, International Foundation for Autonomous Agents and Multiagent Systems Richland, SC, vol. 1, pp. 136–142 (2008)

[13] Porikli, F.: Trajectory distance metric using hidden markov model based representation. In: IEEE European Conference on Computer Vision, PETS Workshop (2004)

[14] Porikli, F., Haga, T.: Event detection by eigenvector decomposition using object and frame features. In: CVPRW 2004: Proceedings of the, Conference on Computer Vision and Pattern Recognition Workshop (CVPRW 2004), vol. 7, p. 114. IEEE Computer Society Press, Los Alamitos (2004)

[15] Reitsma, P.S.A., Pollard, N.S.: Evaluating motion graphs for character navigation. In: Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation, pp. 89–98. Eurographics Association Aire-la-Ville, Switzerland (2004)

[16] Singh, S., Naik, M., Kapadia, M., Faloutsos, P., Reinman, G.: Watch Out! A Framework for Evaluating Steering Behaviors. In: Egges, A., Kamphuis, A., Overmars, M. (eds.) MIG 2008. LNCS, vol. 5277, p. 200. Springer, Heidelberg (2008)

[17] Sud, A., Andersen, E., Curtis, S., Lin, M., Manocha, D.: Realtime path planning for virtual agents in dynamic environments. In: Proc. of IEEE VR (2007)

[18] Thalmann, D., Musse, S.R.: Crowd Simulation. Springer, Heidelberg (2007)

[19] Treuille, A., Cooper, S., Popovic, Z.: Continuum crowds. ACM Trans. Graph. 25(3), 1160–1168 (2006)