Pedestrian Anomaly Detection Using Context-Sensitive Crowd Simulation

Cory D. Boatright   Mubbasir Kapadia   Jennie M. Shapira
Norman I. Badler
University of Pennsylvania
{coryb, mubbasir, jshapira, badler}@seas.upenn.edu

Abstract

Detecting anomalies in crowd movement is an area of considerable interest for surveillance and security applications. The question we address is: What constitutes an anomalous steering choice for an individual in the group? Deviation from "normal" behavior may be defined as a subject making a steering decision the observer would not, provided the same circumstances. Since the number of possible spatial and movement configurations is huge and human steering behavior is adaptive in nature, we adopt a context-sensitive approach to assess individuals rather than assume population-wide homogeneity. When presented with spatial trajectories from processed surveillance data, our system creates a shadow simulation. The simulation then establishes the current, local context for each agent and computes a predicted steering behavior against which the person's actual motion can be statistically compared. We demonstrate the efficacy of our technique with preliminary results using real-world tracking data from the Edinburgh Pedestrian Dataset.

1. Introduction

Anomaly detection is increasingly important in modern security operations, which must observe increasing numbers of people for suspicious behavior. By automating the detection of such behavior, we can lift the burden on personnel and help focus their limited resources. Anomaly detection remains an open research problem because of the challenge in finding a model to serve as the basis of normality while accommodating the diverse range of human behavior. Previous efforts have used such techniques as Gaussian Mixture Models and Hidden Markov Models to define how an average person may act in a particular location with outliers being declared anomalous. A more robust model of "normal" that properly reflects the qualitatively different situations a person may experience is still needed.

Modeling human behavior is precisely the aim of crowd simulation, making these two research endeavors complementary. Data-driven approaches to simulation in particular try to generalize the relationship between environmental stimuli and a corresponding action, making them a strong fit to this application. Training such models on real-world data has presented problems with the unpredictability of what will be observed, and subsequent disagreement of model and human is blamed on the steering algorithm. However, with a high-quality model it is reasonable to question which is truly abnormal. For instance, an intoxicated person's behavior would show that the simulation model is not always at fault. With an adequate simulation, we can analyze the behavior of real people without artificially restricting expectations to averages and other statistical figures.

We propose an anomaly detection system which uses a simulation of "shadow agents" to represent real pedestrians. The system maintains a score for each person according to deviations from their shadow agent's navigation. Our simulation uses a data-driven, compound model of steering which dynamically adjusts each agent's decisions as its environment evolves from its own perspective. The idea of contexts for a crowd are not new, but we extend this idea by allowing each individual to determine its own context rather than setting a crowd-wide context. This model of anomaly detection has several advantages over other techniques. First, the system permits a variety of appropriate behaviors co-existing together rather than assuming the agents are homogeneous. Second, the system guards against the problem where a small, early difference has unnecessarily large influence on the anomaly score by accumulating short-term deviations. This metric depends on the validity of the steering model used, be it our context-sensitive model or any other algorithm. This framework simultaneously checks both the population and the model's accuracy, as an overabundance of anomaly detections are strong evidence of an inaccurate steering algorithm.

This paper makes the following contributions:

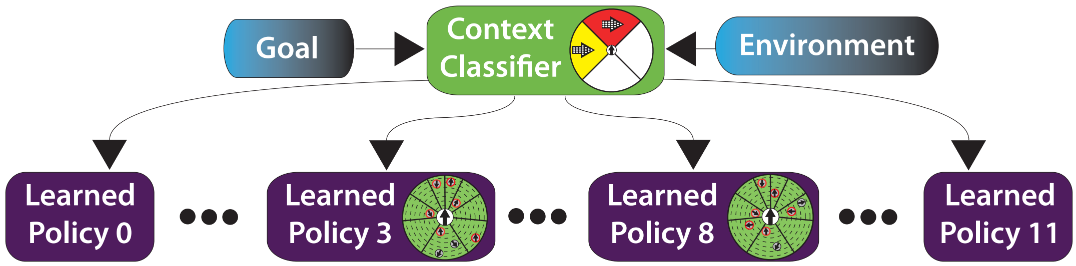


Figure 1. Our compound steering model dynamically chooses between classifiers based on the agent's environment.

- A framework for detecting anomalous pedestrian trajectories in real-time which uses crowd simulation as the basis for comparison and is sensitive to the context each individual is experiencing rather than enforcing a group norm.

- A real-time, cumulative scoring model which is robust against late-starting anomalous behavior, does not artificially weight early decisions higher than those occurring later, and reveals inaccurate models when used on real data.

In Section 2 we frame this paper in the past work found in the literature. Section 3 gives more detail of our simulation, with the anomaly detection discussed in Section 4. Last, we give preliminary results of the technique in Section 5 with conclusions in Section 6.

2. Related Work

This paper proposes to bridge the gap between two areas of research: crowd simulation and anomaly detection in pedestrian movement. While we provide a review of the most applicable crowd literature, those interested in a more thorough survey of the field are directed to [17, 24]. Similarly we give a brief look at some of the common anomaly detection techniques, with further surveys of such work being [6, 4].

Crowd Simulation and Evaluation. Early crowd simulation [19] focused on agent throughput: getting many agents to move on screen and look like a group. In the quarter-century since that seminal work, the field has expanded and moved towards representing more complex dynamics. Emulation of the cognition behind human decision-making [26, 21, 1] has been an active area of research, and provides support for individual roles in the simulation.

In contrast to cognitive approaches, data-driven techniques [15, 11, 13] use machine-learning to map agent stimuli to actions. These techniques seek to fit a single model to the full spectrum of scenarios an agent may encounter through best-match databases. Other works [10, 16, 25] use clustering of their databases to account for the possible encounters which lead to different actions given the same stimuli.

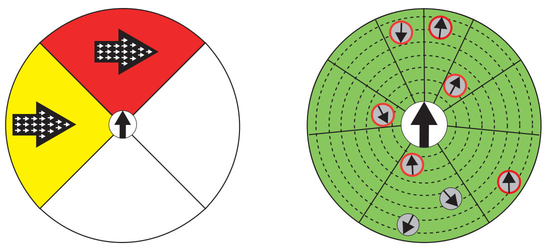
Evaluation of crowds has often been by subjective observation, but statistical techniques have been proposed [7, 22, 8, 12, 9]. We leverage the concept of quantitative crowd metrics for our own anomaly detection system.

Anomaly Detection. In the interest of automated surveillance, computer vision has been interested in a variety of techniques and applications of anomaly detection. The most common technique is to use observations of a real population to fit a model of normal behavior. By focusing on the general flow of the crowds [5], these statistical models can then be used to detect high-level anomalous behavior such as an emergency evacuation [3]. Other works have focused on specific behavior of an individual, but not steering within a crowd [27, 20].

Comparison to the Literature. Both fields have acknowledged the problem of acquiring sufficient real-world data for training models and the potential for synthetic data in developing and training these systems [3, 18, 2]. This work is the realization of such suggestions, as we use an active crowd simulation as the model for normal behavior.

Furthermore, the model itself is egocentric, with each agent in the simulation capable of experiencing a different steering context from its neighbors. This is an extension to [12], where an entire crowd must be considered under the same context. Through the use of steering contexts and a hierarchical data-driven model, we avoid the single-model problem of defining a universally normal behavior for qualitatively different dynamic environments.

3. Hierarchical Steering Model



(a) (b)

Figure 2. The environment is classified using long-horizon density and average trajectory tracked in each region, seen left. A shorter-range, more precise feature set seen right is used by the selected specialized model to decide the agent's next action.

We use a compound machine-learned model for agent steering, outlined in Figure (1). This model is constructed by first identifying qualitatively different steering scenarios an agent may encounter during a simulation, which we call steering contexts. These contexts represent variation such as cross traffic, oncoming traffic, and varying population densities. Each context has a specialized model trained for it, and a top-level classifier is fit to take an agent's environment and decides which context model should be used.

The action space for our model is discretized footsteps [23] and we use synthetic training data from a short-horizon, space-time planner as a steering oracle algorithm. Scenarios representing each context are stochastically generated and the oracle's decisions are recorded. We then use the GPL C5.0 decision tree library (`www.rulequest.com`) to train a model for each foot in each context.

The features used in classifying a context focus on general regional information, particularly each region's population and the average velocity of the agents present. A second feature set is used for more precise measurements of nearby agents. The area around the subject is divided into slices with a higher resolution to the front to simulate human vision. Each slice records the discretized distance to the nearest agent as well as the agent's relative velocity to the subject. Both sets are visualized in Figure (2).

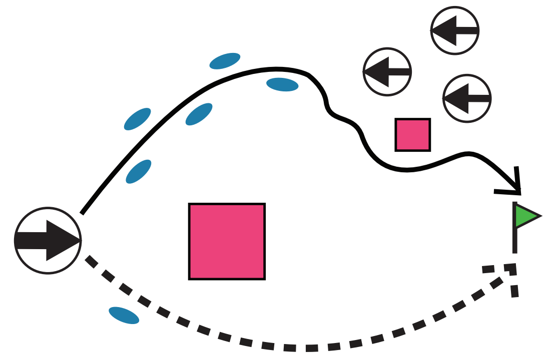


Figure 3. Shadow agents are forced to take the route of the person. After the first step above, the agent and person agree on the subsequent steering choices, reducing the likelihood of an anomaly.

4. Technique

Our system first creates a "shadow" agent in the simulation for each tracked person in the real world. Then we calculate when the divergence between the two is sufficient to merit flagging the behavior as anomalous. Section 4.1 explains how our data-driven model for steering is converted into an observational tool applicable to real humans. The calculation details are given in Section 4.2.

4.1. The Shadow Simulation

Our system takes in tracked data of pedestrians and extracts the necessary information for running a shadow simulation. A shadow agent is created for each person, with the person's first tracked position and last tracked position becoming the agent's spawn and goal points, respectively. The tracking data is also used to force the shadow agent to follow the person's path. Figure (3) illustrates a person's choice to turn left rather than right having large consequences in the total trajectory as more obstacles and people must be avoided to reach the goal. Forcing the agent along the real path instead of simply simulating the scene and comparing the resulting trajectories nullifies inconsequential path diversity. With limited knowledge of each pedestrian's internal state, such singular differences are not sole indicators of anomalies.

At the beginning of each simulated footstep, the agent uses the compound model from Section 3 to project its future expected position. It also compares its current position, which is the end of the previous footstep, against the person's real position. These measurements are used in

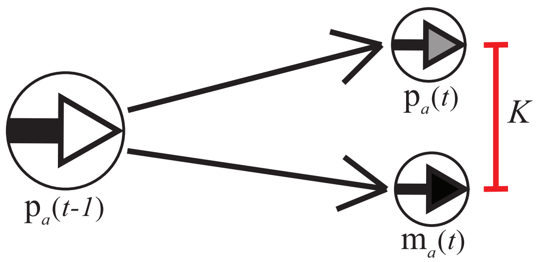


Figure 4. Regular comparison is made between the position of a person and that of its corresponding shadow agent in the virtual world.

Equation (1) to initiate an update of the agent's anomaly score.

Once indicated by a sufficiently high cumulative score, the person is flagged as anomalous by the simulation. This anomaly flag can optionally be removed with enough subsequent expected behavior.

4.2. Flagging Anomalies

At each measurement time t, every agent a has two positions, the real-world position $\mathbf{p}_a(t)$ and the position indicated by the simulation $\mathbf{m}_a(t)$. We use the indicator function in Equation (1) to decide whether or not the deviation from one step to another is significant based on difference kernel K. The tunable parameter d adjusts the sensitivity of the system's detection to allow for such things as measurement error in the tracking data.

$$\mathbf{1}\left(a,t\right)=\begin{cases}1 & \text{if } K\left(\mathbf{p}_a\left(t\right),\mathbf{m}_a\left(t\right)\right)\geq d\\ 0 & \text{else}\end{cases} \tag{1}$$

We let the variable $s_a(t)$ be the score for agent a at time t. The value of $s_a(t)$ is defined according to Equation (2) where ω is a constant decay amount subtracted from the score when normal behavior is observed, χ is the confidence value of the shadow agent's decision from the compound model, and γ is set to reflect the expected accuracy of the specialized classifier used for this particular step. We constrain the value of $s_a(t)$ to be nonnegative.

$$s_a\left(t\right)=\sum_{i=0}^{t}\chi\left(i\right)\gamma\left(i\right)\mathbf{1}\left(a,i\right)-\omega\left(1-\mathbf{1}\left(a,i\right)\right) \tag{2}$$

Tuning ω adjusts the time window over which too many deviations result in higher scores, with larger values creating a more forgiving system. The benefit of this decay-based accumulation function is that an anomaly can start at any time and the score maintained as the shadow agent moves through various contexts. This is an improvement over using a finite time window, where enough early normal behavior can dilute the ability to detect late anomalies through an average score.

We define τ_{anom} to be the score threshold which indicates anomalous behavior in a pedestrian. Additionally, let $\tau_{\text{norm}}\leq\tau_{\text{anom}}$ be a score threshold which indicates a return to normality. The latter is chosen to introduce hysteresis in the detection system to prevent rapid toggling of the anomaly flag. τ parameters can be chosen together with ω to set a desired cooldown time.

Each agent then has a Boolean flag f_a which at time t has the value set by Equation (3).

$$f_a(t)=\begin{cases}1 & \text{if } s_a\left(t\right)\geq\tau_{\text{anom}}\\ f_a\left(t-1\right) & \text{if } \tau_{\text{norm}}<s_a\left(t\right)<\tau_{\text{anom}}\\ 0 & \text{if } s_a\left(t\right)\leq\tau_{\text{norm}}\end{cases} \tag{3}$$

5. Results

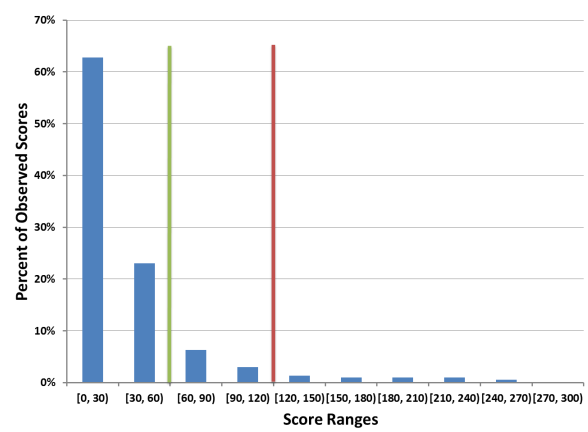
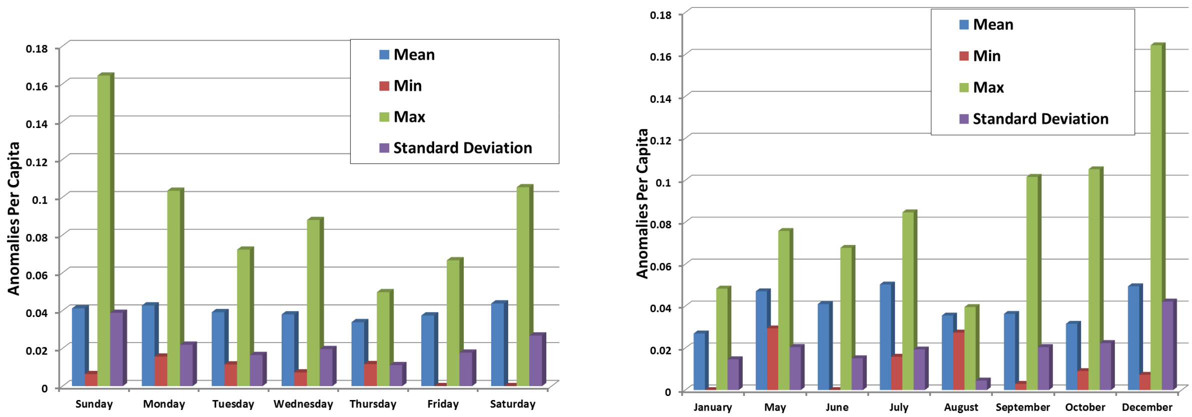


Figure 5. Histogram of score values from running the system used to find values for the anomaly and normality thresholds. The red and green lines are anomaly and normal thresholds, respectively.

To test our system, we used the Edinburgh Informatics Forum Pedestrian Database [14]. Our compound steering model consists of 12 contexts, with each context using 5000 sample scenarios to generate training data. An additional 1000 sample scenarios were withheld for each context as a validation set. The models were evaluated for accuracy using this set to calculate our values for γ, seen

Context Number	0	1	2	3	4	5	6	7	8	9	10	11
γ	.79	.79	.80	.81	.80	.80	.80	.80	.81	.80	.79	.80

Table 1. The accuracy across the specialized classifiers is highly uniform, making no particular context a strength or weakness for the anomaly detection scores.



(a) (b)

Figure 6. Statistical analyses of anomalies per capita for days of the week and months of the year.

in Table (1). A shadow simulation was created for each day of the database, and a histogram of anomaly scores generated using an ω value of 0.1. The distribution of scores can be seen in Figure (5) and strongly suggest the choice of 120 for τ_{anom} and 60 for τ_{norm}, owing to the small value for ω.

Figure (6) shows statistical analyses for the number of anomalies our system detected per capita for each day of the week and month of the year from the dataset. The population count varied greatly for each of the data points, ranging from 5 to 2804. However, the average anomalies per capita across the days and months remained consistent under our system, providing a validation of its robustness. We also note the weekend has a particularly high standard deviation for anomalies detected, indicative of the less uniform crowd flow during those days. Not all months were present in the dataset, and May consisted of only 3 days of tracking information.

Manual inspection of the simulation provided an interesting observation where we noticed anomalous agents under seemingly normal circumstances. On review of the dataset, we found that the floor can reflect the person, causing two agents to be spawned in the same location. In this case the agents continuously try to separate from each other but cannot, causing the high anomaly score.

6. Conclusions and Future Work

This paper presented an initial exploration into the use of a data-driven, context-sensitive crowd simulator for pedestrian anomaly detection. We used our prototype framework to examine the Edinburgh Dataset by reporting the computed anomalies for the tracked pedestrian trajectories over 115 days.

We are actively exploring several avenues of future work. Our framework is fast, operating on a day of tracked data in minutes, suggesting potential for use in live surveillance. Our system is currently constrained to pedestrian movement, but we would also like to expand the contexts we use to include such things as small groups walking together to increase the quality of our algorithm and the breadth of its impact. Correlation-based metrics are another set of scoring techniques we could explore. An important validation of our technique will be to compare it against existing anomaly detection frameworks, such as the model provided with the dataset [14].

7. Acknowledgements

The research reported in this document/presentation was performed in connection with Contract Number W911NF-10-2-0016 with the U.S. Army Research Laboratory. The views and conclusions contained in this document/presentation are those of the authors and should not be interpreted as presenting the official policies or position, either expressed or implied, of the U.S. Army Research Laboratory, or the U.S. Government unless so designated by other authorized documents. Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation heron.

We also thank Corey Novich for her assistance with our illustrations.

References

[1] J. M. Allbeck. CAROSA: a tool for authoring NPCs. In *MIG 2010*, pages 182–193. Springer, 2010.

[2] J. M. Allbeck and N. I. Badler. Distributed Video Sensor Networks. In B. Bhanu, C. V. Ravishankar, A. K. Roy-Chowdhury, H. Aghajan, and D. Terzopoulos, editors, *Distributed Video Sensor Networks*, pages 193–205. Springer London, London, 2011.

[3] E. L. Andrade, S. Blunsden, and R. B. Fisher. Modelling Crowd Scenes for Event Detection. *ICPR '06*, pages 175–178, 2006.

[4] H. M. Dee and S. A. Velastin. How close are we to solving the problem of automated visual surveillance? *Machine Vision and Applications*, 19:329–343, May 2007.

[5] L. F. Henderson. The statistics of crowd fluids. *Nature*, 229:381–383, 1971.

[6] W. Hu, T. Tan, L. Wang, and S. Maybank. A survey on visual surveillance of object motion and behaviors. *IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews*, 34(3):334–352, 2004.

[7] M. Kapadia, S. Singh, B. Allen, G. Reinman, and P. Faloutsos. Steerbug: An Interactive Framework for Specifying and Detecting Steering Behaviors. In *SCA 2009*, volume 1, pages 209–216, 2009.

[8] M. Kapadia, S. Singh, and W. Hewlett. Egocentric affordance fields in pedestrian steering. *I3D '09*, 1(212):215–224, 2009.

[9] M. Kapadia, M. Wang, S. Singh, G. Reinman, and P. Faloutsos. Scenario space: Characterizing coverage, quality, and failure of steering algorithms. In *SCA 2011*, pages 53–62. ACM, 2011.

[10] K. H. Lee, M. G. Choi, Q. Hong, and J. Lee. Group behavior from video: a data-driven approach to crowd simulation. In *SCA 2007*, volume 1, pages 109–118. Eurographics Association, 2007.

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

[12] A. Lerner, Y. Chrysanthou, A. Shamir, and D. Cohen-Or. Context-Dependent Crowd Evaluation. *Computer Graphics Forum*, 29(7):2197–2206, 2010.

[13] A. Lerner, E. Fitusi, Y. Chrysanthou, and D. Cohen-Or. Fitting behaviors to pedestrian simulations. In *SCA 2009*, volume 1, pages 199–208, New York, New York, USA, 2009. ACM Press.

[14] B. Majecka. *Statistical models of pedestrian behaviour in the Forum Master of Science Artificial Intelligence School of Informatics University of Edinburgh*. Masters, University of Edinburgh, 2009.

[15] R. A. Metoyer and J. K. Hodgins. Reactive pedestrian path following from examples. In *16th International Conference on Computer Animation and Social Agents*, volume 20, pages 149–156, Nov. 2003.

[16] S. R. Musse, C. R. Jung, J. C. S. Jacques Jr., and A. Braun. Using computer vision to simulate the motion of virtual agents. *Computer Animation and Virtual Worlds*, 18:83–93, 2007.

[17] N. Pelechano, J. Allbeck, and N. Badler. *Virtual Crowds: Methods, Simulation, and Control*. Morgan & Claypool, 2008.

[18] F. Z. Qureshi and D. Terzopoulos. Surveillance camera scheduling: A virtual vision approach. *Multimedia systems*, 12:269–283, 2006.

[19] C. W. Reynolds. Flocks, herds and schools: A distributed behavioral model. *ACM SIGGRAPH Computer Graphics*, 21(4):25–34, Aug. 1987.

[20] P. Scovanner and M. F. Tappen. Learning pedestrian dynamics from the real world. In *2009 IEEE 12th International Conference on Computer Vision*, pages 381–388. Ieee, Sept. 2009.

[21] W. Shao and D. Terzopoulos. Autonomous pedestrians. *Graphical Models*, 69(5-6):246–274, Sept. 2007.

[22] S. Singh, M. Kapadia, P. Faloutsos, and G. Reinman. SteerBench: a benchmark suite for evaluating steering behaviors. *Computer Animation and Virtual Worlds*, 20:533–548, 2009.

[23] S. Singh, M. Kapadia, G. Reinman, and P. Faloutsos. Footstep navigation for dynamic crowds. *Computer Animation and Virtual Worlds*, 22:151–158, 2011.

[24] D. Thalmann and S. R. Musse. *Crowd Simulation*. Number May. Springer, 2007.

[25] P. Torrens, X. Li, and W. A. Griffin. Building Agent-Based Walking Models by Machine-Learning on Diverse Databases of Space-Time Trajectory Samples. *Transactions in GIS*, 15:67–94, July 2011.

[26] Q. Yu and D. Terzopoulos. A decision network framework for the behavioral animation of virtual humans. In *SCA 2007*, pages 119–128. Eurographics Association, 2007.

[27] H. Zhong, J. Shi, and M. Visontai. Detecting unusual activity in video. *CVPR 2004*, 2:819–826, 2004.