Crowds in Context: Evaluating the Perceptual Plausibility of Pedestrian Orientations

Christopher Peters, Cathy Ennis, Rachel McDonnell and Carol O'Sullivan

[1] Graphics Vision and Visualisation Group, Trinity College, Dublin

Abstract

We describe a work-in-progress evaluating the plausibility of pedestrian orientations. While many studies have focused on creating accurate or fast crowd simulation models for populating virtual cities or other environments, little is known about how humans perceive the characteristics of generated scenes. Our initial study, reported here, consists of an evaluation based on static imagery reconstructed from annotated photographs, where the orientations of individuals have been modified. An important focus in our research is the consideration of the effects of the context of the scene on the evaluation, in terms of nearby individuals, objects and the constraints of the walking zone. This work could prove significant for improving and informing the creation of computer graphics pedestrian models. Our particular aim is to inform level-of-detail models.

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Computer Graphics]: Perception

1. Introduction

Modelling pedestrian behaviour has been subject to a great deal of investigation and many noteworthy achievements have been made in synthesising high-level behaviours (see [ST05] [PAB07] for some recent examples). Yet studies focusing on how viewers perceive and evaluate the plausibility of the resulting properties and behaviours remain scant. This issue is of great significance in the domain of graphics, where accuracy of simulation may be traded for performance in order to display larger crowds than would otherwise be possible while minimising perceived errors in the behaviour. In this paper, we focus on the first step in an evaluation methodology that we are developing to allow us to better elucidate this complex issue. This first step is described in Section 3, and considers the construction of experiments to evaluate peoples general impressions of static scenes, capable of accounting for factors such as the density, orientation and positioning of pedestrians: in this paper, we focus primarily on orientation. Rather than considering pedestrian properties in isolation, a central theme in our methodology is to also consider the *context* in which it appears. There can be many different aspects relating to context, which we classify here as being of two general types: firstly, nearby pedestrians and objects that may affect an individual, and secondly, the type of walking zone that an individual inhabits. Based on this methodology, the results and implications of a perceptual study we conducted are described in Section 4. Our results demonstrate the importance of considering the role of context in users' evaluation of the plausibility of pedestrian properties. We conclude by describing future work in Section 5, where we intend to extend the methodology to account for dynamic aspects of pedestrian behaviour, and broaden those aspects we consider as relating to context.

Figure 1: *Photograph of the open zone (left) and the virtual reconstruction (right). The orientations of individuals in these reconstructions were modified according to rules for comparison through perception studies.*

2. Related Work

Two previous approaches are similar to our work, particularly in terms of data collection and annotation. In Lerner et al. [LCD07], a database is constructed from input video. Pedestrians in the video are manually tracked to generate a set of trajectories, which are stored as examples in the database. At runtime, the database is queried for similar examples matching those of the simulated pedestrian and the closest matching example is selected as his trajectory. Lee et al. [LCHL07] recorded crowd videos in a controlled environment from an aerial view. Users must manually annotate video frames with static environment features and can semi-automatically track multiple individuals in order to provide their trajectories. This data informs an agent movement model to provide a crowd that behaves similarly to those observed in the video. Although both cases involve manual annotation of crowd behaviour, both use the data as exemplars for generating behaviours rather than as a basis for conducting perception studies.

3. Evaluation Methodology

Our evaluation methodology consists of four phases. The first three phases detail the collection (Section 3.1), annotation (Section 3.2) and reconstruction (Section 3.3) of a virtual scene closely approximating the original, in terms of pedestrian positions and orientations (see Figures 1 and 2 for examples and an overview of the process). Modifiers are then applied to the reconstruction (Section 3.4) to create artificial formations that can be compared with those relating to the real scene.

3.1. Data Collection Phase

A number of photographs were taken of two locations representative of two prototypically opposite types of pedestrian movement zones. We classify these prototypes as *unconstrained* or *open* zones (see Figure 1) and *constrained* or *corridor* zones (see Figure 2(a)). An open zone typically has many exits and entrances, through which pedestrians will be crossing in many varying directions. In contrast, a corridor zone typically places great constraints on the movement of the pedestrians (through physical or social means) and will tend to enforce bi-directional movement.

3.2. Annotation Phase

Each photograph was annotated manually (see Figure 2(b)) in order to highlight groupings and their orientations. Groups are deemed to consist of one or more individuals. Each group is designated by an ellipse, which expands to cover all members of the group. Each ellipse is colour coded according to whether the corresponding group is static (*red*) or mobile (*yellow*). All orientations are classified as belonging to one of the following 8 rotations specifying cardinal directions: 0, 45, 90, 135, 180, 225, 270 and 315. Each direction is associated with a unique colour code, to aid recognition of the general characteristics of the scene.

Figure 2: *Reconstructing the scene. An (a) initial photograph is (b) annotated with groups and their orientations. Camera parameters are matched up with those of the original camera so (c) characters can be placed corresponding to the transformations of the real people resulting in (d) a virtual scene with a similar composition to the real one.*

3.3. Reconstruction Phase

Once a photograph has been annotated, we reconstruct the scene by using it as a view-port background in *3D Studio Max* and fitting our 3D model by manually tweaking the virtual camera parameters to ensure an acceptable fit between the photograph and the model. Next, the positions of virtual characters are manually matched up with their real-life counterparts from the photograph, providing a good approximation to the composition of the original scene (see Figure 2(c)). Orientations are classified as belonging to one of the 8 cardinal directions and all pedestrians are orientated to match the nearest corresponding direction from this list.

3.4. Orientation Modification Rules

We have enumerated a number of basic rules for modifying pedestrian orientation and accounting for context. The orientations of all individuals in the scene is set to one or more of the 8 cardinal directions, according to one of the four following basic orientation rules:

1. Original - the orientation of each individual is aligned to match the annotated orientation from the original scene (Figure 3(a)).
2. Random - each individual is assigned an orientation chosen at random from one of the 8 cardinal directions (Figure 3(b)).
3. Uniform - a single orientation is chosen at random from one of the 8 directions and all characters are aligned to match it (Figure 3(c)).
4. Even - individuals are chosen at random and assigned an orientation from one of the 8 cardinal directions so as to fit into an overall even distribution (Figure 3(d)).

3.4.1. Contextual Rules

In addition to these four basic orientation rules, the methodology includes three context sensitive rules for altering orientations. These take into account other individuals and the environment:

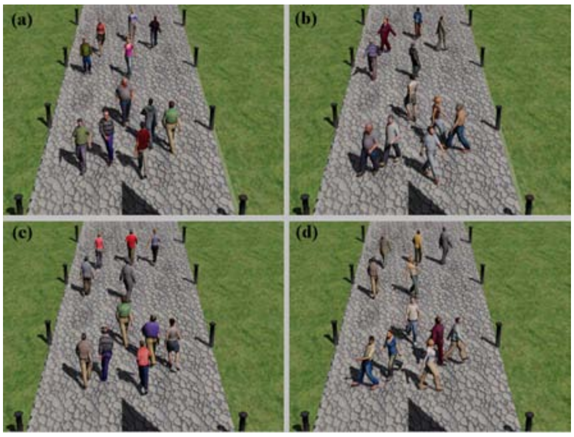

Figure 3: *Example of the basic orientation rules. Reconstruction (a) of the original photograph, (b) random rule, (c) uniform rule and (d) even rule applied to all pedestrians.*

1. Group sensitive - the orientations are decided on a group rather than an individual basis.
2. Flow sensitive - the orientation of each individual is chosen randomly from a subset of the 8 cardinal directions. This subset is created from the allowable flow directions for the position of the character, based on a ground flow-tile representation.
3. Adjacency sensitive - the orientation of each individual is chosen randomly from a subset of the 8 cardinal directions. This subset is created by disqualifying those directions which result in the character facing towards inappropriate adjacent zones, e.g. individuals appearing to walk into each other.

When combining rules, basic rules are applied first and then contextual rules are applied afterwards, overriding the basic rules. In the case of flow sensitivity for example, the result is that the basic rule is only applied in ground areas that have not been assigned flow directions.

4. Perceptual Study

We conducted a perceptual experiment with the purpose of testing the methodology described in Section 3, and as a baseline for developing future studies for more complex pedestrian properties. This experiment focused on testing orientation and context rules for characters in static scenes, in order to determine their effect on perceived plausibility.

4.1. Design

We defined a number of *master scenes*, each of which was a reconstruction based on an original, annotated photograph of the scene (see Section 3.3). In these scenes, characters' positions have been matched as closely as possible to those in the original photograph, and orientations are matched according to the closest of the 8 cardinal directions defined. No attempt was made to match the actual models or poses of characters with the original scenes; however, we ensured that the model and poses assigned to a character differed between each image. Fabricated scenes were created manually by applying our orientation rules (see Section 3.4) to the master scenes in order to change the orientations of the individuals. There were therefore three general categories of imagery used in our study: master scenes corresponding to the real scenes, images derived from the master scenes where orientation was altered as defined by the non-contextual orientation rules, and images derived from the master scenes where orientations were defined according to the context sensitive rules. For the context sensitive rules, flow lanes were created manually for the scene from an inspection of our annotated corpus of pedestrian movement. Adjacency rules were not applied, as the crowd density in our corpus did not create orientation conflicts e.g. pedestrians appearing to walk into each other.

4.2. Goals and Hypotheses

The goals of our perceptual study were to test if participants would find the reconstructed real scenes more plausible than those where the orientations were synthetically generated, and for the latter, to compare responses between those that were created with and without the use of the context sensitive rules. Our first hypothesis was that participants would preferentially rate those original images reconstructed from actual scenes (the master scenes) above those created manually according to the application of our rules. Secondly, we hypothesised that for the images where the orientations were synthetically generated, those that considered the contextual rules detailed in Section 3.4.1 would receive more favourable ratings, than those not considering any form of context.

4.3. Method

Twenty five participants (9F 16M) ages 18 to 30, sat in front of a projected display. Divided into 2 groups, they were shown randomised orderings of images and given an instruction sheet: two photographs of the corridor and open zone were shown and they were told that the images they were about to see were derived from real photographs, but in some the character orientations were real, while in others they were synthetically generated. For each image displayed, participants were asked if they thought the orientations of the characters were real or synthetically generated. Between each trial, a blank-screen was displayed for 5 seconds after which a noise alerted participants for the next trial.

A total of 72 images were displayed for 10 seconds each, 36 of which were master scenes where the orientations matched an original photograph, and 36 of which were orientations modified by the rules in Section 3.4. An image could thus be categorised as belonging to one of the following seven different types depending on the rules applied to the orientations: Original, Random no context (RandomNC), Random with context (RandomC), Even no context (EvenNC), Even with context (EvenC), Uniform no context (UniformNC), and Uniform with context (UniformC). Those listed as *With Context* employ a basic rule modified by the contextual rules (Section 3.4.1), while those listed as *No Context* employ only a basic rule.

4.4. Results and Discussion

We averaged responses over each of the three repetitions for each orientation type. A one factor ANalysis Of VAriance (ANOVA) with repeated measures showed a main effect of Orientation type ($F_{2,48} = 42.256, p < 0.001$). Post-hoc analysis was then performed using a standard Newman-Keuls test for pairwise comparisons among means.

We found that the original scenes with "Real" orientations were judged as *real* significantly more times than either the context or no-context virtual scenes ($p < 0.02$ in all cases), implying that people are able to distinguish the real cases from the synthetic ones based primarily on differences in orientation. Also, the virtual context scenes were judged as real significantly more times than the virtual no-context scenes ($p < 0.02$) implying that our context rules increased plausibility of orientations.

We then expressed the responses for all virtual scenes as offsets from the average "Real" values for the open and corridor scenes for each orientation type. We performed a two factor ANOVA with repeated measures and the results are shown in figure 4. There were main effects of Context ($F_{6,144} = 17.1, p < 0.001$) and orientation types ($F_{6,144} = 6.6, p < 0.003$), but no main effect of location implying that overall, there was no difference in ratings between open and corridor scenes. There were also two way interactions between location, orientation and context ($p < 0.007$ in all cases).

When looking at these interactions, it can be seen that the context rules performed well in the corridor scenes, but had mixed effects in the open scenes. Looking at the former, the context rules clearly improved the realism of virtual orientations. The orientation type that was perceived to be closest to Real orientation was RandomC, with its average rating only 2.67% less than the the average rating of the master scenes. The Uniform orientation performed almost as well, with a rating of 4% less than the master scenes. For all orientations in the corridor scene, with context scenes performed significantly better than no context scenes ($p < 0.004$ in all cases). The success of the context rules in the corridor scene suggests that the rules used in this study would be sufficient for setting the orientations of crowds in this type of area.

For the open location, both UniformNC and RandomNC were the best performing orientations, performing significantly better than their corresponding scenes with the context rules ($p < 0.001$ in both cases). However, the Even orientation with context rules performed significantly better than the no context orientation ($p < 0.009$). The poor performance of UniformC and RandomC rules could be due to the flow context rules not covering enough of the scene area to have a significant effect on the orientations. It could also be due to some obstacle-adjacency issues, where the models in the scene appear to be walking into or facing towards obstacles. This mostly occurred near the centre of the image, where the viewers' focus was concentrated. The addition of more flow lanes and adjacency rules could rectify the poor

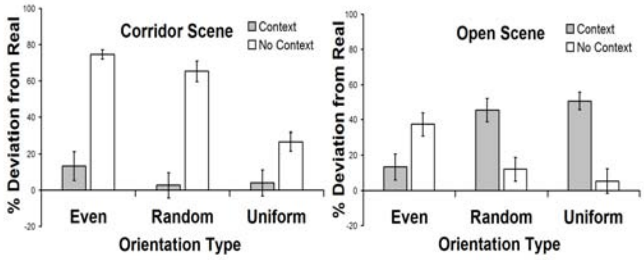

Figure 4: *Deviation from average Real responses for Corridor (left) and Open (right) scenes for each orientation with and without context.*

performance of the context rules for an unconstrained zone for these orientations. Future eye-tracking studies may provide more clues to resolve this, and could be of relevance for employing level of detail schemes.

Overall, it can be seen that the success of the application of the contextual rules is, to a greater or lesser degree, contingent on the type of zone that the pedestrians inhabit. In the corridor zone, the orientation rules are more significant in contributing towards the perception of plausibility. However, for the open zone, these rules do not have such a significant effect, with both context and non-context rules performing well for different orientation types. For these type of zones, additional flow lanes and obstacle-adjacency rules could be implemented to improve realism.

5. Future Work

In addition to the orientation rules, we have also defined similar rule-sets for the positioning of characters and will be conducting perceptual studies for these in the future. Another important area of future work is to consider dynamic scenes and the dynamic properties of pedestrian behaviours. This will be challenging, as the manual annotation of scenes is a cumbersome, error-prone and time-consuming process: an investigation of semi-automatic methods for aiding this process would be of much benefit, similar to those recently adopted by other authors [LCHL07].

References

[LCD07] LERNER A., CHRYSANTHOU Y., DANI L.: Crowds by example. *Computer Graphics Forum 26*, 3 (2007), 655–664.

[LCHL07] LEE K. H., CHOI M. G., HONG Q., LEE J.: Group behavior from video: a data-driven approach to crowd simulation. In *SCA '07: ACM SIGGRAPH/Eurographics symposium on Computer animation* (2007), pp. 109–118.

[PAB07] PELECHANO N., ALLBECK J. M., BADLER N. I.: Controlling individual agents in high-density crowd simulation. In *SCA '07: ACM SIGGRAPH/Eurographics symposium on Computer animation* (2007), pp. 99–108.

[ST05] SHAO W., TERZOPOULOS D.: Autonomous pedestrians. In *SCA '05: ACM SIGGRAPH/Eurographics symposium on Computer animation* (2005), pp. 19–28.