Perceptual Effects of Scene Context and Viewpoint for Virtual Pedestrian Crowds

CATHY ENNIS, CHRISTOPHER PETERS, and CAROL O'SULLIVAN, Trinity College Dublin

In this article, we evaluate the effects of position, orientation, and camera viewpoint on the plausibility of pedestrian formations. In a set of three perceptual studies, we investigated how humans perceive characteristics of virtual crowds in static scenes reconstructed from annotated still images, where the orientations and positions of the individuals have been modified. We found that by applying rules based on the contextual information of the scene, we improved the perceived realism of the crowd formations when compared to random formations. We also examined the effect of camera viewpoint on the plausibility of virtual pedestrian scenes, and we found that an eye-level viewpoint is more effective for disguising random behaviors, while a canonical viewpoint results in these behaviors being perceived as less realistic than an isometric or top-down viewpoint. Results from these studies can help in the creation of virtual crowds, such as computer graphics pedestrian models or architectural scenes, and identify situations when users' perception is less accurate.

Categories and Subject Descriptors: I.3.7 [**Computer Graphics**]: Three-Dimensional Graphics and Realism—Virtual reality

General Terms: Experimentation, Human Factors

Additional Key Words and Phrases: Perception, virtual crowd formation

ACM Reference Format:

Ennis, C., Peters, C., and O'Sullivan, C. 2011. Perceptual Effects of Scene Context and Viewpoint for Virtual Pedestrian Crowds. ACM Trans. Appl. Percept. 8, 2, Article 10 (January 2011), 22 pages.
DOI = 10.1145/1870076.1870078 http://doi.acm.org/10.1145/1870076.1870078

1. INTRODUCTION

Humans are generally expert at recognizing and rating the behaviors related to other individuals. Given our high exposure to human behavior in our social environment, coupled with internal mechanisms, such as mirror neurons [Gallese et al. 1996] and those related to mentalizing and theory of mind [Premack and Woodruff 1978], we can discern the behavior of others at multiple levels of sophistication to varying degrees of detail and certainty. Mirror neurons fire when we observe someone performing an action or when we perform a similar action ourselves. They have been linked to cognitive functions such as empathy and the ability to understand the intentions of others. Theory of mind is an innate ability humans possess to predict or explain behavior of others by attributing goals, desires, or emotions to their actions. These capabilities are applicable too to robots and computer characters with a humanoid appearance [Schilbach et al. 2006], in which cases a viewer may rate the realism of

This work was funded by the Science Foundation Ireland.
Author's address: C. Ennis, Graphics Vision and Visualisation Group, Trinity College Dublin, Ireland; email: ennisca@cs.tcd.ie.

DOI 10.1145/1870076.1870078 http://doi.acm.org/10.1145/1870076.1870078

an artificial motion or behavior favorably or unfavorably even if they are at a loss to identify precisely why it appears so. What is less clear, however, and becoming evermore important as computer processing capabilities allow the display and simulation of increasingly large numbers of humanoid agents, is the question of the degree to which humans can discern and rate features of crowds of humans or humanoid figures.

We investigate the effects of individuals' positions and orientations (in isolation and in combination with each other) on the viewer's perception of realism. We also examine the possible effects of variation across different scenes on the perceived plausibility of a scene. Finally, the effects of camera viewpoint on the perceived realism of such crowd scenes is explored. In addition, we present a tool to semi-automate the image generation process by incorporating contextual rules for the various formations, which helps to eliminate manual errors and speeds up the process of creating the corpus to be used in the experiment.

We hypothesize that the context in which each individual is perceived is one of the most important factors affecting the plausibility of a simulated crowd. By context, we refer to each individual's relationship with respect to both its environment and neighbors. These relationships should be appropriate and consistent with common experience: Pedestrians tend to be seen walking on paths, are often directed toward exit or goal positions when mobile, do not walk into obstacles or other individuals, and members of groups are usually in proximity to each other and may have other properties in common. In order to test this, we compare scenes where individuals' positions and orientations have been generated according to contextual rules with those directly derived from real scenes, and those that have been randomly generated. It is important to remember that the objective of this study was not to physically replicate real scenes but to generate formations of pedestrians in a way that is perceptually real to the viewer by considering the context of the scene. This work is particularly important for evaluating and modeling pedestrian and crowd behavior (see Shao and Terzopoulos [2005] and Pelechano et al. [2007] for recent examples).

After reviewing related work in Section 2, Section 3 outlines the methodology for constructing the scenes used in our experiments. We conducted three perceptual experiments, which are described in Sections 4, 5, and 6. Finally, we present our results and discuss possible applications in Section 7.

2. RELATED WORK

There are many methods of simulating crowds and crowd behavior, such as social force models [Helbing and Molnár 1995], path planning [Lamarche and Donikian 2004], and behavioral models incorporating perception and learning [Shao and Terzopoulos 2005], or sociological effects [Musse and Thalmann 1997].

In recent years, taking human perception of virtual worlds, their characters and associated animations into account has been increasingly recognized as an important factor in achieving more realistic scenes. While much research has been conducted into issues such as the perception of animation and motion of individuals (e.g., Reitsma and Pollard [2003] and McDonnell et al. [2007]), or spatial awareness (e.g., Henry and Furness [1993]) very little is known about human perception of crowds, in particular crowd formation.

In previous research, we investigated pedestrian crowds in an urban context [Peters et al. 2008]. We used static scenes consisting of crowds of humanoid characters to study the effects of orientation on users' perception of realism. The orientations of some characters were matched with those of their counterparts from real scenes, while others were artificially modified. We found that participants were consistently able to distinguish between the scenes containing the real orientations and those with artificial orientations. In addition, orientation rule types that accounted for the context of the scene were judged overall to be more realistic than those utilizing naïve rule sets. This initial indication that humans seem adept at judging such details raises intriguing questions as to what aspects are contributing to their impressions of realism and to what degree. We replicated these results in a further study [Ennis et al. 2008] and found similar effects for pedestrian scenes that contained position and orientation information.

There are two previous approaches that are similar to our work, particularly in terms of data collection and annotation. Lerner et al. [2007] set out to generate crowds that displayed varied behaviors individually without defining an explicit behavior model. They used a data-driven example-based approach to achieve this, allowing the agents to learn from real-world examples. A database was constructed from input video of real-world pedestrian behaviors by manually tracking pedestrians in the video to generate a set of trajectories, which were stored as examples in the database. At runtime, the database was queried for similar examples that match those of the simulated pedestrian, and the closest matching example was selected as the resulting trajectory.

Lee et al. [2007] used a data-driven approach to simulate virtual human crowds imitating real crowd behavior. They recorded crowd videos in a controlled environment from an aerial view. Users manually annotated video frames with static environment features and then could semiautomatically track multiple individuals in order to provide their trajectories. This data informed an agent movement model to provide a crowd that behaved similarly to those observed in the video. Although both cases involved manual annotation of crowd behavior, both used the data as exemplars for generating behaviors rather than as a basis for conducting perception studies.

It has been shown that viewpoint can effect the perception of facial expressions. Lyons et al. [2000] found that the angle a facial mask was tilted at altered how happy or sad participants judged the facial expression of the mask to be. They also recreated these results using a human face. Bülthoff et al. [1995] also found that recognition of computer-generated 3D objects is viewpoint dependent in that participants performed better when looking at the objects through familiar viewpoints. Similar results were also found by Tarr [1995], where participants performed better in recognizing a formation of blocks when presented from a familiar viewpoint. When the object was rotated, reaction times were slower. This implies that familiarity with objects is developed through specific viewpoints and that we depend on this for object recognition and identifying facial expressions for virtual objects and faces. Little is known about how virtual crowds and their surroundings are perceived, however, nor whether our perception of a virtual pedestrian crowd will be affected as the viewpoints change. We take the first step in examining this in relation to behaviors of virtual crowds by studying whether the perceived realism of virtual formations changes across viewpoints.

3. METHODOLOGY

Our methodology consists of four phases. The first three phases refer to the collection (Section 3.1), annotation (Section 3.2), and reconstruction (Section 3.3) of virtual scenes closely approximating original still images in terms of pedestrian positions and orientations. The final phase is the modification (Section 3.4) of aspects of these scenes according to rules, in order to produce artificial formations to be compared with each other and the real reconstructions.

3.1 Data Collection Phase

A number of videos were taken of two different locations, each representing an archetypical pedestrian movement zone. We refer to these as *constrained* or *corridor* locations and *unconstrained* or *open* locations. An open location represents a relatively large space where pedestrians tend to be seen crossing in many varying directions due to the presence of a multitude of possible exits and entrances. In

Fig. 1. Reconstructing the scene. An (a) initial still image is (b) annotated with groups and their orientations. Camera parameters are matched up with those of the original camera so that (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.

contrast, a corridor location is more constrained, usually with a single entrance/exit at either end and, therefore, tends to enforce bidirectional movement.

A number of still images were extracted from each video, to be used as a basis from which to create reconstructions of the scenes depicting the real positions and orientations of individuals. These will be referred to here as the *real* category of scenes. To minimize the variation in responses from participants, the density of pedestrians in the two locations (corridor and open) was kept as equal as possible. Based on the area of the zone visible to the viewer, it was estimated that 30 pedestrians in the open zone corresponds in crowd density terms to 12 in the corridor zone, and, therefore, still images with these numbers were selected from the extracted video stills for the respective zone types.

3.2 Annotation Phase

Each still image was annotated manually to highlight individuals' positions and orientations and their groupings, if any (see Figure 1(b)). For the purpose of this article, a group is described as a pedestrian unit of one or more individuals and is designated according to their localization in space and aided by a visual inspection of the video clip surrounding the still image being annotated. Each group was designated by an ellipse, which covered all members of the group and was color-coded according to whether the corresponding group was static (*black*) or mobile (*yellow*). The orientations of individuals were classified as belonging to one of the following eight rotations specifying cardinal directions: 0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315°. Each direction was associated with a unique color code, to aid

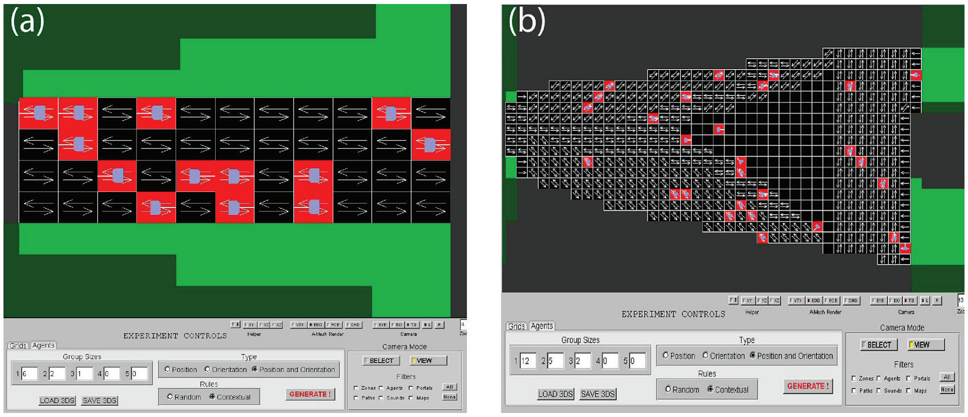

Fig. 2. View of our in-house tool for generating pedestrian transformations based on contextual rules. These images show the corresponding grids used (a) for the corridor location and (b) for the open location images.

visual recognition of the general characteristics of the scene, such as the number of groups containing one, two, or three individuals.

3.3 Reconstruction Phase

The reconstruction phase consists of recreating virtual replicas of the real images captured and annotated in the previous phases. Once a still image has been annotated, we reconstruct the scene by using it as a viewport background in 3D Studio Max and fitting our 3D model by manually tweaking the virtual camera parameters to ensure an acceptable fit between the still image and the model. Next, the positions of virtual characters are manually matched up with their real-life counterparts from the still image, providing a good approximation to the composition of the original scene.

These reconstructions were replicas in the sense that they matched certain aspects of the real scenes, such as individuals' positions, orientations and groupings, whereas we did not attempt to replicate individuals' appearance, clothes, poses, and gender.

3.4 Modification Phase

We investigated pedestrian formations according to which of the following three aspects would be studied: only individuals' positions, only their orientations, and both their positions and orientations. For each, we generated the required positions and/or orientations semi-automatically, using a combination of our own stand-alone tool (see Figure 2) and 3D Studio Max. Our tool was important not only for reducing the workload in creating the scenes to be used in the experiment, but also for reducing placement errors and aiding in the replicability of the experiment.

Rather than looking at the pedestrian characteristics in isolation, since an important consideration for us is the *context* of individuals in the scene, a number of context rules were defined for positioning and orientating characters. There can be many different aspects relating to context, which belong to three general types:

(1) nearby pedestrians, objects and obstacles that may affect an individual,
(2) the type of walking area that an individual inhabits, for example, in order to specify the general direction of *flow* in that area [Chenney 2004] and
(3) group properties that play a role in people's perceptions of crowds and pedestrians (e.g., group size and the number of groups in a scene).

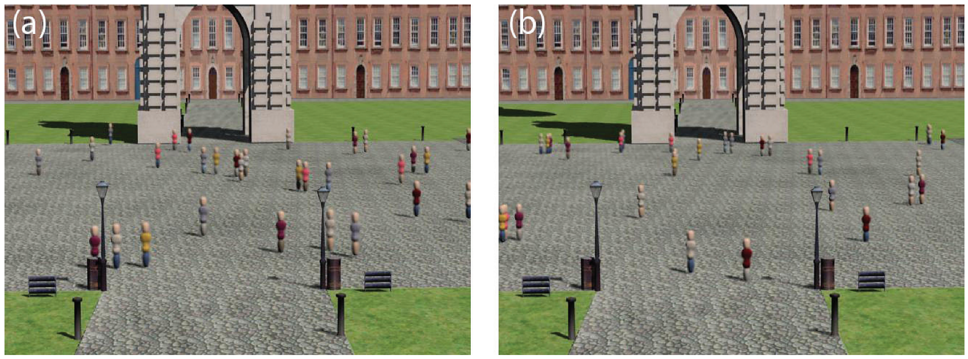

Fig. 3. Position images for the open scene: (a) the virtual representation of the real still image and (b) the open scene with position context rules.

For those rules that involve modifying the original positions or orientations, a number of steps must be taken in our tool to allow for the automatic generation of data. A grid is created in order to fit the area that will appear in the final scene. Each cell in the grid is then manually assigned attributes, such as walkability and flow direction(s), if any. This process only needs to be conducted once for each area from which renderings must be conducted: in our study, this process needed to be conducted only twice: once for the open location and once for the corridor location. One can then select a position type and a rule to apply—clicking on a button will then generate the resulting transformations according to the rules selected.

After a scene has been generated with the tool, it is exported to 3D Studio Max as a set of dummy nodes, each of which contains the transformation for a particular pedestrian. Each node is manually associated with a mesh, either a posed human figure or else a directionless pawn figure, the latter of which is used in the position studies. The rules and steps involved in the modification of the scenes are described in the following sections.

3.4.1 *Position.* For the images modified to create the position formations, the characters are displayed as a *pawn* figure with no discernible orientation. The images used for the position block of the experiment are shown in Figure 3. Here, the only information available is the position of the characters, and this is modified in the following ways.

(i) Real. The position of each individual in the scene is the same as the positions of the pedestrians in the still image.
(ii) Random. Each individual is assigned a random position on the grid.
(iii) Context-Based. Each individual is assigned a random position that adheres to our context rules, which are listed in the following text.

Position Context Rules

(i) Bounds Sensitive. An individual can only be assigned a random position that is part of a designated walkable area. In these experiments, grass was regarded as being out-of-bounds during the application of contextual rules.
(ii) Group Sensitive. Individuals will be assigned a position to maintain the number and size of groups in the original still image. The positions of the individuals within the group will be assigned a random formation from a set of appropriate formations and will not be derived from the real image. The template formations vary in number depending on group size (e.g., four for groups

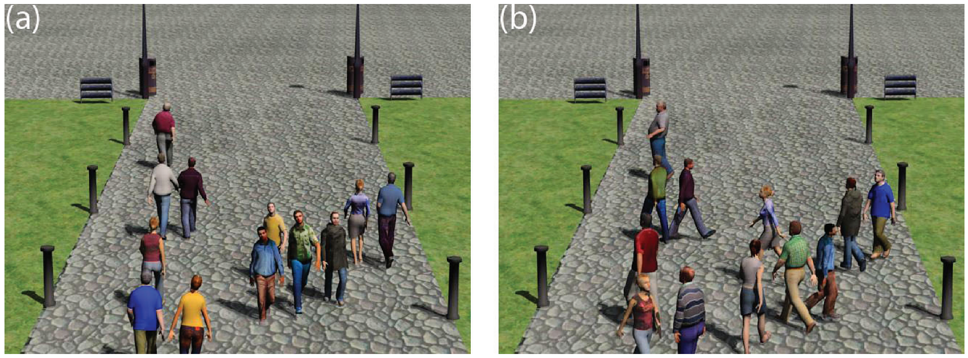

Fig. 4. Orientation images for the corridor location: (a) the virtual representation of the real still image and (b) the corridor scene with orientation random rules.

with two agents, nine for groups of three agents), but agents are placed in a group in adjacent cells facing the same direction.

3.4.2 *Orientation.* The modified images used for the orientation block contained standard human character models with discernible orientations (see Figure 4). The character positions remained the same as those from the original still image, while only their orientations were changed.

(1) Real. The orientation of each individual in the scene is the same as the orientations of the pedestrians in the still image.
(2) Random. Each individual is assigned one of the eight cardinal orientations on a random basis.
(3) Context. Each individual is first assigned a random orientation, which then is altered to obey our orientation context rules (listed in the following text).

The orientation context rule can be specified by the following three rules when determining the orientation of each individual.

(i) Flow Sensitive. The orientation of each individual is chosen randomly from a subset of the eight cardinal directions. This subset is created from the allowable flow directions for the position of the character, based on a ground flow-tile representation.
(ii) Adjacency Sensitive. The orientation of each individual is chosen randomly from the eight cardinal directions, but any direction leading to inappropriate facings is disqualified (e.g., a character walking into a lamp post or another character would be considered an inappropriate facing). The grid in our tool can be altered to allow different distances from obstacles (as can be seen in Figure 2). For these experiments, we used a distance of one grid tile from the obstacle before the area could be assigned appropriate flow lanes, but this distance could have been any number of grid tiles wide. All of our agents were considered to be dynamic agents, rather than stationary, so our adjacency rules do not consider agents conversing face to face but will allow agents to appear to follow one another.
(iii) Group Sensitive. Each individual within a group of two or more is assigned the same orientation, rather than on an individual basis. Since our agents were all dynamic, agents were not orientated facing toward each other

In order to resolve conflicts between the rules, the rules were applied in the order provided earlier: flow sensitivity, then adjacency sensitivity and, if applicable, group sensitivity. Members of a group

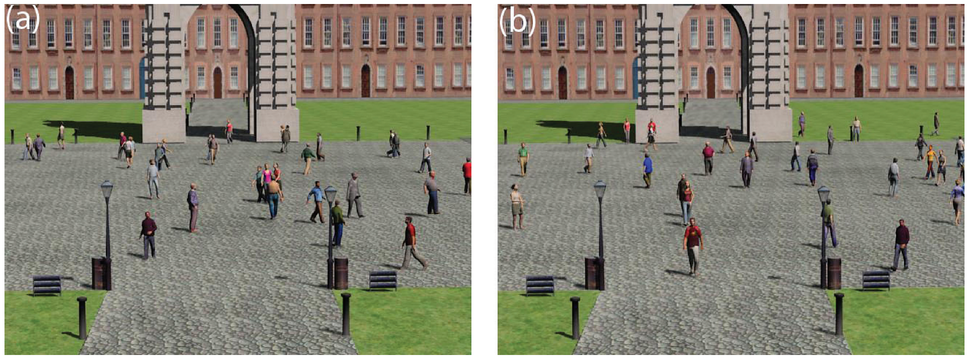

Fig. 5. Position and orientation images for the open scene: (a) position context rules with random orientation and (b) both position and context rules.

were all orientated in the same direction, based on the direction in which most of the group members were facing in accordance with the previously applied rules. The scenes containing orientation were constructed manually.

3.4.3 *Orientation and Position.* For the images modified for the orientation and position formations, the scene is again divided into a grid. As in Section 3.4.2, this block of the experiment used normal human characters with discernible orientations. The context rule images used for this block are shown in Figure 5. Both the orientation and position will be modified in the following ways.

(1) Real. The position and orientation of each individual in the scene is the same as the positions and orientations of the pedestrians in the still image. Different still images were used for the generation of real scenes for each block of our experiment.
(2) Random. Each individual is assigned a random position on the grid and an orientation from the eight cardinal orientations.
(3) Position Context (Context Pos). Each individual is assigned a random orientation with a position according to the position context rules as explained in Section 3.4.1.
(4) Orientation Context (Context Ori). Each individual is assigned a random position on the grid and an orientation according to the orientation context rules as explained in Section 3.4.2.
(5) Orientation and Position Context (Context Both). Each individual is assigned an orientation and position that obey both our orientation and position context rules.

4. EXPERIMENT 1: POSITION AND ORIENTATION

The first perceptual experiment we conducted was to see how realistic the rules explained in the previous section were perceived to be.

4.1 Procedure

Thirty-two participants (12F, 20M) age 18 to 30, were seated in front of a computer screen. They were told that the experiment consists of three blocks and were given an instruction sheet: two photographs of the corridor and open locations were shown and participants were told that the images they were about to see were derived from real photographs. However, in some, the character formations were real, while in others, they were synthetically generated. For the first block of the experiment, the participants were told to focus only on the positions of the characters. For each image displayed, participants

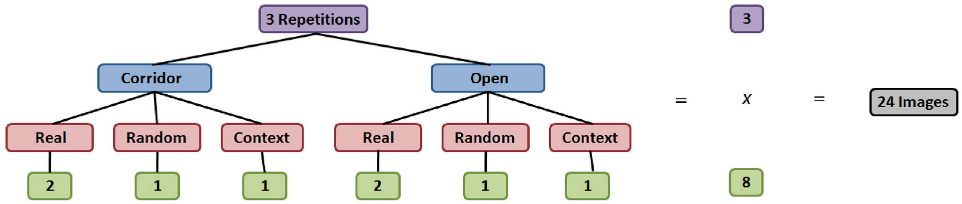

Fig. 6. Breakdown of images used for Blocks 1 and 2 of Experiment 1, which examine position only and orientation only, respectively.

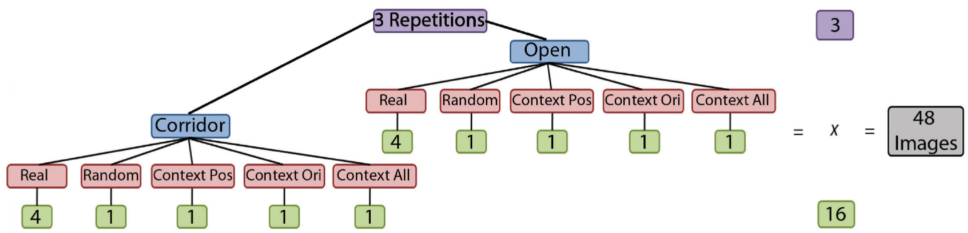

Fig. 7. Breakdown of images used for Block 3 of Experiment 1, looking at position and orientation.

were asked if they thought the positions of the pawn figures were real or synthetically generated. For the second block, participants were asked to look at the orientations of the characters only and judge if they were real or synthetically generated. For the final block of the experiment, participants were asked to take both position and orientation of the characters into account and to judge whether the scenes were real or synthetically generated. The reason that we presented the blocks in this order was to avoid biasing participants. If the pawn figures were viewed after the humanoid characters, this could have caused them to perceive the scenes as less realistic due to the reduced realism of the characters, which was not the effect being tested. Furthermore, the scenes with position and orientation combined were presented during the final block, to prevent participants from taking position into consideration when conducting the orientation only trial. Between each trial, a blank-screen was displayed for 5 seconds, after which the number of the next trial was displayed alerting participants.

For the first block of the experiment, a total of 24 images were displayed for 4 seconds each, 12 of which were master scenes where the positions matched an original still image, and 12 of which were positions modified by the rules in Section 3.4.1. An image could thus be categorized as belonging to one of the following three different types: Real, Random, and Context. Each experiment block contained a unique Real still image, so for Experiment 1, a total of three different real images were used (one for each experiment block). See Figure 6 for details of the images used in this experiment.

For the orientation experiment block, again 24 images were shown for 4 seconds each, 12 containing scenes with real orientations and 12 containing orientations modified as described in Section 3.4.2 with Real, Random, and Context categories. Examples of images used in this experiment can be found in Figure 4.

For the final experiment block, looking at position and orientation, a total of 48 images were displayed for 4 seconds each. Of these, 24 contained positions and orientations matching still images of real scenes, and 24 images were modified according to the rules mentioned in Section 3.4.3. For this block of the experiment, an image could be categorized as belonging to one of the following five different types: Real, Random, Position Context, Orientation Context, and Both Context. The breakdown of the images used for this experiment can be seen in Figure 7.

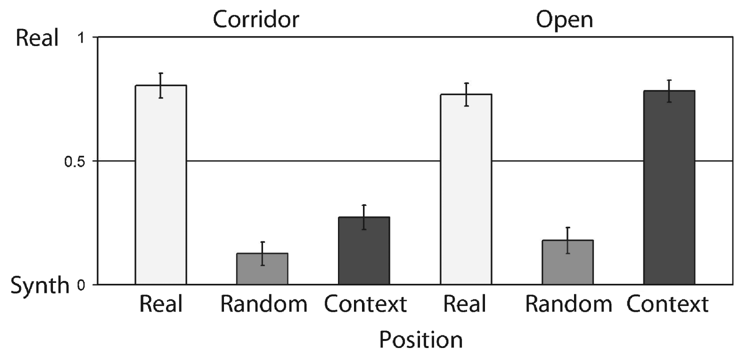

Fig. 8. Graphs showing position results for Corridor and Open locations.

4.2 Results

We found no significant differences based on the ordering of the images within the experiment blocks in participants' responses, implying that there were no ordering effects within the individual experiments.

4.2.1 *Position Results.* We averaged responses over each of the three repetitions for each position type. A two factor ANOVA with repeated measures showed a main effect of Scene ($F(1, 31) = 17.895$, $p < 0.01$), in that the open location was perceived to be more realistic than the corridor location, and position type ($F(2, 62) = 70.077$, $p < 0.01$), where the real positions were perceived to be more real than virtual positions. There was also an interaction between the two ($F(2, 62) = 23.476$, $p < 0.01$), where the real positions were perceived as real more often for the corridor location than the open location, whereas the virtual positions were perceived as real more often for the open location. Posthoc analysis was then performed using a standard Newman-Keuls test for pairwise comparisons among means.

We found that the original corridor scenes with real positions were judged as real significantly more times than either the context or random virtual corridor scenes ($p < 0.01$ in all cases), implying that participants are able to distinguish the real cases from the synthetic ones based primarily on differences in position for constrained zones. However, participants perceived the corridor scenes with context-based positions to be more real than those with random positions ($p < 0.05$).

Looking at the open location, while participants perceived the real positions to be more real than the random positions ($p < 0.01$), they judged the scenes with context-based positions to be as realistic as those with the real positions. Figure 8 shows these findings on a scale of 0 to 1, where 0 means they were perceived as synthetically generated and 1 means they were perceived as real.

As shown in Figure 8, while the context rules applied did improve the perceived realism of the scene over random positioning, averaging across locations, the participants could still distinguish the real positions from the synthetically generated ones. For the open location, the context rules applied had a greater effect on participants perception of the scene. The fact that participants judged the scenes with our context rules to be as real as those scenes with real positions suggests that these rules could be an adequate way to populate scenes when positioning characters in an open or unconstrained location. While the rules do not have such a strong effect on the realism of corridor scenes, it has been shown here that they could be an adequate method to initialize a pedestrian crowd scene. While perhaps not as effective as manually placing characters in appropriate positions, our method is less time consuming and provides a suitable alternative to random positioning. While this is interesting in

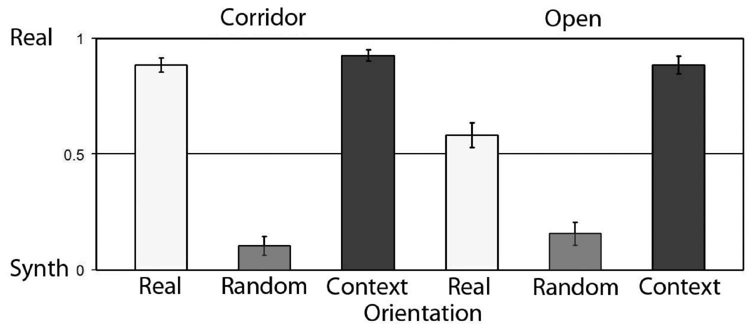

Fig. 9. Graphs showing orientation results for Corridor and Open Scenes.

itself, the effects of the positioning of characters are not very useful for practical applications without taking into account the orientations of characters.

4.2.2 *Orientation Results.* We averaged responses over each of the three repetitions for each orientation type. A two factor ANOVA with repeated measures showed a main effect of location ($F(1, 31) = 11.508$, $p < 0.01$), where the corridor location was perceived to be more real for this experiment, and orientation type ($F(2, 62) = 162.04$, $p < 0.01$) where context orientations were perceived to be the most real, followed by the real orientations, with the random orientations being judged the least real. There was also an interaction between location and orientation ($F(2, 62) = 12.040$, $p < 0.01$), where participants judged the real and context scenes as real more often for the corridor location but judged the random scene as real more often for the open location. Posthoc analysis was then performed using a standard Newman-Keuls test for pairwise comparisons among means.

We found that the original corridor scenes with real orientations were judged as real significantly more times than the random virtual corridor scenes ($p < 0.01$), but the corridor scenes with context rules were judged to be as real as the original scenes, implying that participants are unable to distinguish the real cases from orientations generated using our context rules for constrained zones.

Looking at the open location (see Figure 9), while participants perceived both the real and context-based orientations to be more real than the random orientations ($p < 0.01$), they judged the scenes with context-based orientations to be more realistic than those with the real orientations. One possible explanation for this unexpected result is that the unconstrained nature of the open scene provided less contextual cues with which to judge the validity of the characters' orientations. This would suggest that there may be more perceptual tolerance when viewing character formations in these scene types than in more constrained areas, where viewers are more adept at spotting peculiarities. Another possible reason for this was that the high level of realism of the context scenes had an effect on participants' perception of the real scenes.

As illustrated in Figure 9, the results show that the context rules greatly affect how real the orientations are perceived to be. In our previous study [Peters et al. 2008], our results showed similar trends. However, the addition of the adjacency sensitive rules to the context scenes resulted in the participants perceiving the orientations as significantly more realistic than in our previous study, in particular for the open scene. It is known that fixations in a scene are task dependent [Yarbus and Haigh 1973], so for this experiment, participants would likely be scanning each character in the scene to determine how realistic their position and orientation appears to be. Because of the camera viewpoint (similar to the ones in our previous study), the closest characters appear near the bottom of the image. This region of the grid in our tool was hugely affected by the addition of the adjacency rules since this area joins an out-of-bounds area. It is possible that the participants focus on where the characters are most salient,

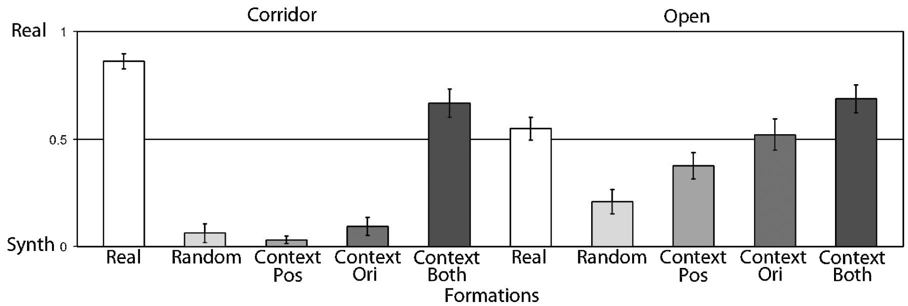

Fig. 10. Graphs showing position and orientation results for Corridor and Open Scenes.

for the duration of exposure to the image, so this could be a possible explanation for the positive effect on perceived realism caused by the addition of the adjacency rules. As with the results for the position rules, the combination of position and orientation rules are where the most useful information is in terms of practical applications for placing human characters in virtual scenes.

4.2.3 *Position & Orientation Results.* We averaged responses over each of the three repetitions for each type of formation. A two-factor ANOVA with repeated measures showed a main effect of location ($F(1, 31) = 15.754, p < 0.01$), where the open location was perceived to be more real than the corridor location, and formation type ($F(4, 124) = 54.093, p < 0.01$), where scenes with both position and orientation context rules were judged to be almost as real as the real scenes. There was also an interaction between both ($F(4, 124) = 16.615, p < 0.01$), where participants perceived the real formations to be more real for the corridor location, but judged the synthetic formations to be more real for the open location. Posthoc analysis was then performed using a standard Newman-Keuls test for pairwise comparisons among means.

We found that for the corridor location, the real formations were judged as real significantly more times than any of the four virtual formations ($p < 0.01$ in all cases). This implies that, given position and orientation information for a constrained location, participants are able to tell real scenes from virtual scenes. Nevertheless, the scenes where both position and orientation context rules had been applied were perceived as real significantly more often than either the random, position context, or orientation context scenes ($p < 0.01$ in all cases).

In the open location, the results were again more complicated, with the real scenes being judged as real more often than the random formations and scenes with position context information only ($p < 0.01$ in all cases). For the scenes with orientation context and both position and orientation context, the participants judged them to be as real as the real scenes. This reinforces the hypothesis that participants find it more difficult to distinguish between real and synthetic formations for less constrained scenes. Figure 10 shows that when context is considered when placing characters, participants' perception of the realism of the characters is greatly improved compared to random placement. While participants can still differentiate between real and synthetic for the corridor scene, they cannot differentiate between them for the open location. Interestingly, for the corridor location, the use of either position or orientation context on its own was not effective, as participants perceived these to be as synthetically generated as the random scenes. However, for the open location, participants perceived the scenes with random positioning and orientation context rules to be as real as both the real scenes and the scenes with both position and orientation context rules applied. This would imply that, in an unconstrained location with a large number of people, orientation seems to be of greater importance than position when it comes to plausibility.

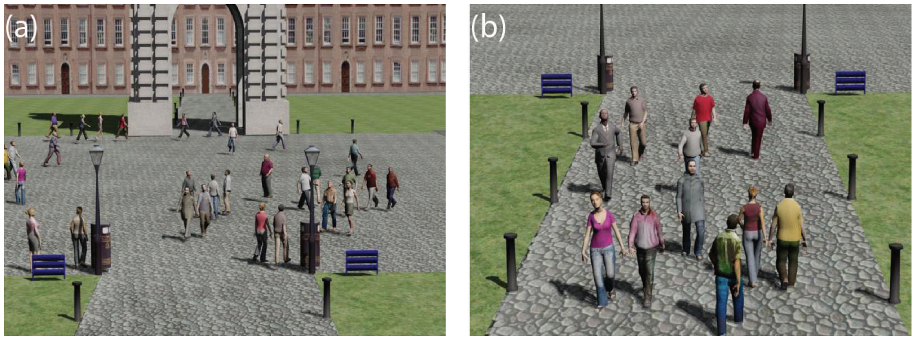

Fig. 11. Showing example of real scenes used in Experiment 2 for (a) the open scene and (b) the corridor scene.

For the orientation-only and orientation-position combined experiment blocks (blocks 2 and 3), participants gave the real images for the open location unexpectedly low ratings (58% and 55%, respectively). For both of these settings, participants rated the images with context-rules as real more often (89% and 69%, respectively). This could have been due to the fact that context rules do in fact look more realistic than real scenes, or it may have been caused by the photos used to make the real scene for these particular blocks containing anomalies that were not present in the other real images. In order to investigate whether these low ratings were a result of the effect of the context images, or the real images themselves, we conducted a second experiment looking at variation across different real scenes, as well as context and random scenes.

5. EXPERIMENT 2: REAL SCENE VARIATION

For each block of Experiment 1, different real formations were used, with participants perceiving some of these as plausible a relatively low number of times. We conducted a second experiment in order to investigate whether any of these results could have been caused by any specific real formation appearing unrealistic.

5.1 Method

Using the same method as outlined in Section 3.3, we created four different real scenes for both the open location and the corridor location. We kept the number of characters in each scene constant, to ensure no effect of population density. There were three different variations of each scene, where we placed different character models in the same positions and orientations. This was done to avoid participants becoming familiar with a specific character model thus influencing their judgements. An example of one of the four different real scenes used for this experiment for each location can be found in Figure 11.

In addition to the real scenes, we created 6 scenes using the context rules as outlined in Section 3.4 for both locations. We also generated 6 scenes for each location using random placement. Unlike Experiment 1, where random placement included out-of-bounds regions such as grass areas, we only allowed random placement within walkable areas. This was due to the difference in the ratio of walkable areas to out-of-bounds regions across the two different locations. For the open location, there is only a small area out-of-bounds toward the back of the scene, whereas in the corridor location, almost half of the scene is designated as out-of-bounds. This difference could have had an effect on participants' perception of random behaviors, since an inappropriate position in the constrained location could be more easily noticed than in the unconstrained location. An example of a context scene for the open

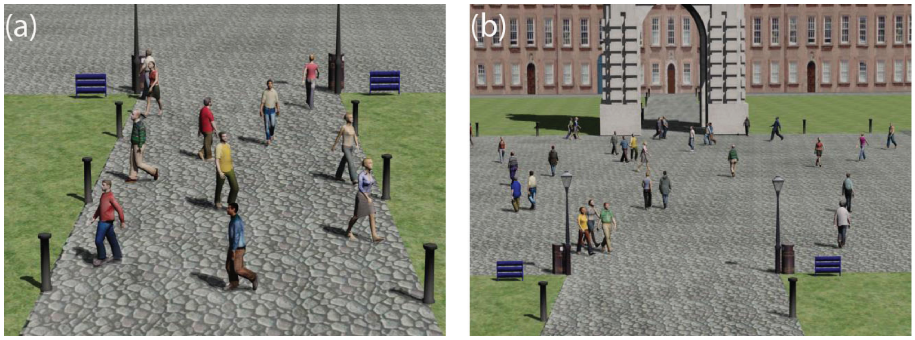

Fig. 12. Showing (a) random and (b) context random scenes for the constrained and unconstrained locations, respectively.

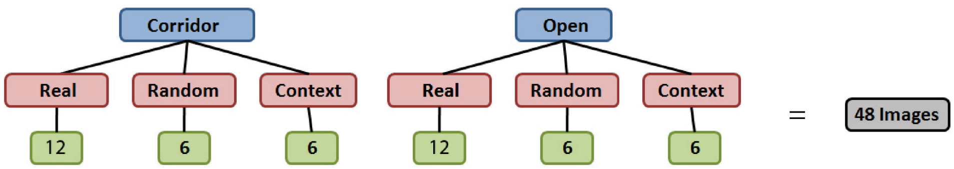

Fig. 13. Breakdown of images used for Experiment 2, looking the effect of different real scenes.

location and a random scene for the corridor location is shown in Figure 12. For each random and context scene, we kept the number of characters constant, with 30 characters in the open location scenes and 11 in the corridor location scenes. For the context scenes, we kept the number and sizes of groups the same as in the real scenes.

5.2 Procedure

Ten participants (1F, 9M), aged 18 to 30, were seated in front of a computer screen and were given an instruction sheet. For each image displayed, participants were asked if they thought the pedestrian formations were from a real scene, or whether they were synthetically generated. A total of 48 images (details can be seen in Figure 13) were displayed for 4 seconds each. Between each trial, a blank screen was displayed for 5 seconds, after which the number of the next trial was displayed to alert participants.

5.3 Results

The images were presented to participants in two different randomized orders. We found no effects of ordering between the two groups.

We averaged responses over each of the three repetitions for each of the four real scenes. There appears some slight variation in the ratings for the real scenes, as shown in Figure 14). In particular, Real 1 appears to have received a lower real rating than the other real scenes for the open location. Looking at this scene, it was the only real scene that contained no groups of three characters. Each other scene contained a group with three characters. The slight difference in clustering of characters in this scene could be one possible reason for the slightly lower real rating for this image. However, an ANalysis Of VAriance (ANOVA) showed there were no statistically significant differences between the four real scenes for either the corridor or open locations. This implies that the low ratings for some of the real scenes in Experiment 1 were not a result of a particular scene appearing any more or less realistic than the others, but rather as a result of the effect of the context and random rules on the

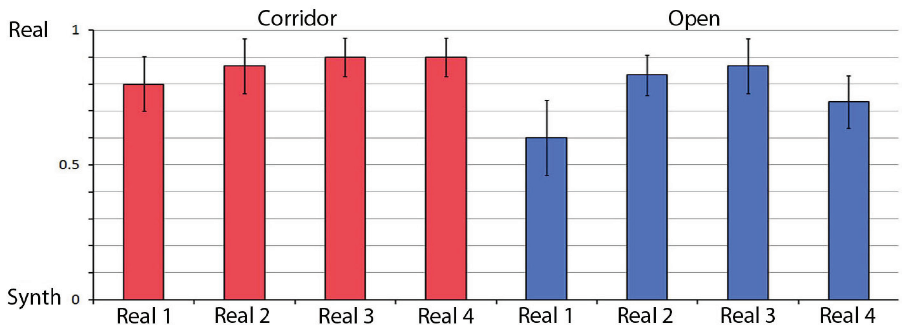

Fig. 14. Graphs showing results for corridor and open locations of four real scenes.

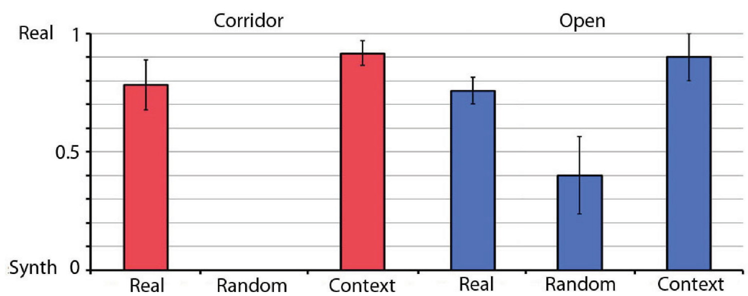

Fig. 15. Graphs showing all results for Corridor and Open Locations (Note: for the corridor location, random scenes received a zero rating averaging across all participants).

participants' perception of the scene. Because of this, we were able to average across all of the real scenes for further analysis.

We averaged responses over each of the three repetitions (where the characters were varied for each repetition, but not the positions and orientations) for each of the three formation types (real, random and context). A one-factor ANOVA with repeated measures showed a main effect of formation type ($F(2, 18) = 45.044, p < 0.0001$). Posthoc analysis was then performed using a standard Newman-Keuls test for pairwise comparisons among means. We found that participants perceived the random formation to be real less often than the context or real formations but that participants perceived the context scenes to be real as often as the real scenes (see Figure 15). This result was expected for the open location, based on the results from our previous experiments, where participants found context scenes to be as realistic as real scenes in many situations. There was a slightly different result for the corridor location from our previous experiment however. Before, participants could distinguish between the real scenes and our context scenes, whereas in this experiment they could not.

We did not find an effect of location type, meaning that the participants' perception of the scenes did not vary between the open and corridor locations. This differs from our results found in Experiment 1, where we found an effect of scene in each block. A possible reason for this result is the alteration of the random placement rules to exclude placement of characters in out-of-bounds regions of the scene. It is likely that the large difference in size of the out-of-bounds areas in the two different locations had an effect on the participants' perception of the scene. However, since they were still able to correctly identify the random placement scenes, it would imply that this was not the only cue used to identify such behaviors. Other contextual information is still important, such as grouping and orientation.

Another factor that could possibly have an effect on a viewers perception of a scene is the viewpoint through which they are viewing the scene and the distance from the camera to the characters. In Experiments 1 and 2, the distance from the characters varies across the locations, and this may have an effect on the perception of the formations within the scenes. Because of the location used to obtain the videos for the real scenes, there is also a difference in the viewpoint between the corridor and open locations. We conducted a third experiment to examine whether the camera viewpoint has an effect on perceived realism in pedestrian crowd scenes.

6. EXPERIMENT 3: CAMERA VIEWPOINT

In our first experiment, we investigated the effects our context rules for pedestrian formations had on participants' perception of realism in two prototypical locations when compared against real and random pedestrian formations. However, it is not clear how generalizable our results are for different viewpoints of a single scene. This is an important factor in most real-time crowd systems, as the user will be navigating through the environment changing the vertical and horizontal positions of the camera. This will mean that a single scene would be viewed through a number of different viewpoints. We need to ensure that, as the viewer navigates through an environment, their perceived sense of realism will not be affected. As a first step to find out how such factors can affect the plausibility of crowd scenes, we conducted a third experiment to examine whether there was any effect of camera viewpoint on the realism of crowd formations.

6.1 Camera Views

We chose four different camera viewpoints, which can be seen in Figure 16. The first viewpoint we chose was eye-level (viewpoint 1). This was due to the prevalence of first person video games, where the player sees everything through the viewpoint of the main character in the game.

The second viewpoint we used was a canonical viewpoint (viewpoint 2). The idea of a canonical viewpoint was introduced by Palmer et al. [1981] and has been shown to aid peoples' memory of virtual objects [Gomez and Shutter 2008]. A canonical viewpoint of an object can be regarded as the viewpoint first imagined visually, or the viewpoint that is selected as the best angle at which to take a photograph. In free exploration tasks, the canonical viewpoint will often be inspected for the longest period of time [Blanz et al. 1999]. As mentioned in Section 2, object recognition has been shown to be viewpoint dependent and in recognition experiments, canonical viewpoints tend to have the lowest response time and error rate. A canonical viewpoint of an object is often a rotation of 10 degrees about each axis [Tarr 1995]. We approximated this viewpoint in our scene by using a slightly elevated camera angle with a small rotation.

The isometric view (viewpoint 3) was chosen as a midpoint between our canonical and top-down viewpoints, as there was a very large difference between these two angles. We positioned the camera higher up and more angled to the scene than our canonical view, but maintained a perspective view of the scene.

Finally, we used a top-down view (viewpoint 4), which can be an important viewpoint to consider when larger crowds and more areas of a scene need to be within view (e.g., in movies with large-scale scenes).

6.2 Scene Construction

In contrast to Experiments 1 and 2, for this experiment, we constructed all our scenes using Metropolis, our crowd system. This had the advantage of allowing us to conduct our experiments on a real crowd rendering system, where our results will eventually be translated and reapplied as rules for the behavior of our pedestrian crowd. Another advantage of using Metropolis is that the environment

Fig. 16. Showing (1) the eye-level and (2) canonical viewpoints for the corridor location with random positioning; and the (3) isometric and (4) top-down viewpoints for the open location with context positioning.

model was more realistic than the model we had used in 3D Studio Max for Experiments 1 and 2, giving participants a heightened sense of realism during the experiment.

In order to create our images for our experiment stimuli, we used our tool to position and orientate the characters in the scene as in Experiment 1. For this experiment, we only included virtual formations, thus creating scenes according to random positioning and according to our full context rules (both position and orientation rules combined). Since we were no longer using 3D Studio Max, it would not have been feasible to exactly match the positions of the characters in a real scene. We wanted to eliminate the distance from the characters to the camera as a possible effect on our results, so we ensured that for both locations the distance between the closest position on the grid to the camera was the same. In order to do this, we placed the camera for the corridor location facing the opposite direction to that in Experiments 1 and 2. Placing the camera at the same position as before would have resulted in it being placed inside a building and the view of the scene being obstructed. Other factors we wished to control were the number of characters in the scene, and the area occupied by the characters. In some viewpoints there are some areas of unoccupied space. This is a result of our confinement of the pedestrians to an area that was visible in each camera viewpoint in order to keep crowd size constant. Therefore, since the eye-level viewpoint covers less of the area than any of the other views, the canonical, isometric, and top-down views all contain areas where there are no pedestrians.

6.3 Procedure

Twenty-three participants (4F, 19M), aged 18 to 30, were seated in front of a computer screen and were given an instruction sheet. For each image displayed, participants were asked if they thought

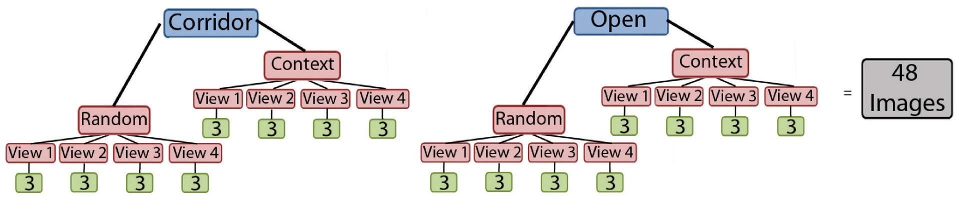

Fig. 17. Breakdown of images used for Experiment 3, looking the effect of different camera viewpoint.

the scenes were realistic or not. A total of 48 images were displayed for 4 seconds each. Details of the images used in this experiment are shown in Figure 17. Between each trial, a blank screen was displayed for 5 seconds, after which the number of the next trial was displayed to alert participants.

6.4 Results

We averaged responses over each of the three repetitions for each type of formation. A two-factor ANOVA with repeated measures showed a main effect of formation type ($F(1, 23) = 25.887$, $p < 0.01$), where scenes with context rules were judged to be more real than the random scenes. There was also a main effect of viewpoint ($F(3, 69) = 4.2014$, $p < 0.01$), where participants judged the canonical view of the scene to be less real than any of the other three views. Interactions were found between location and formation ($F(1, 23) = 7.4572$, $p < 0.05$), where context formations were more real in the corridor location, as opposed to the open location, where random formations were found to be more real. We found a similar effect in our previous study [Peters et al. 2008], where random formations were more plausible in an open location than in a corridor location. This could be explained by the nature of the scenes themselves (i.e., an open location is less restrictive and has more possible flow directions than a corridor location). The final interaction was between formation and viewpoint ($F(3, 69) = 3.1268$, $p < 0.05$), where the top-down viewpoint was perceived to be most realistic for context scenes, whereas the eye-level view was perceived to be most real in the random scenes.

Breaking the analysis down across formation types, by analyzing the participants' responses to context and random scenes separately, we found that for the context scenes, there was no effect of either location or viewpoint. This result indicates that participants judged the context rules applied to a scene to be equally realistic, regardless of whether it is a corridor or open scene and which viewpoint is used. This is an interesting result for us as it implies that our context rules will work effectively using any viewpoint, and, therefore, transitioning between the viewpoints when traversing the virtual city in our crowd system will not result in any loss of perceived realism.

When looking at the responses for the random formations alone, there was a main effect of location ($F(1, 23) = 7.9331$, $p < 0.01$), where the open scene was seen as more real than the corridor scene, as in Experiment 1. This could possibly be due to the fact that the original random positioning rules were used in this experiment. When an altered random positioning was used in Experiment 2 to remove placement in out-of-bounds areas, there was no effect of viewpoint. There was also a main effect of viewpoint ($F(3, 69) = 4.5257$, $p < 0.01$), where the canonical viewpoint was perceived to be the least realistic out of the four viewpoints. One possible explanation for this is that when we look at a scene through a canonical viewpoint, it is almost an ideal viewpoint, containing more scene information than other viewpoints, especially eye-level and top-down. With more information being displayed, anomalies are more easily spotted and the scene will look more unrealistic than for another viewpoint, where any randomness may be disguised by occlusion of characters. Looking at Figure 18, it can be seen that when the camera is at eye-level (viewpoint 1), random formations seem to be more realistic than at any other viewpoint. This result is important for us in terms of using computational savings when

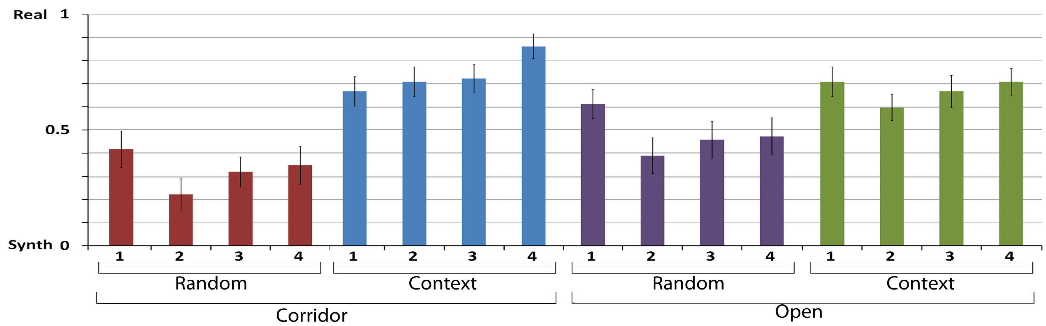

Fig. 18. Shows responses for both locations (Corridor/Open), both formations (Random/Context) and each viewpoint (1-4).

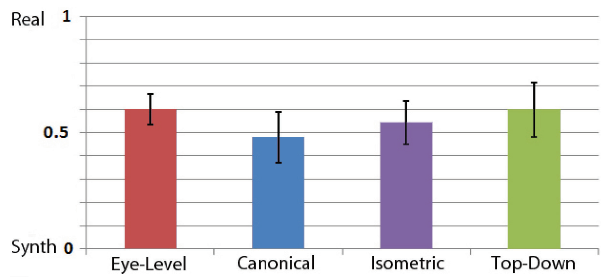

Fig. 19. Shows participant responses for each viewpoint across both locations and both formations.

our system is using this camera viewpoint only, since viewers are less inclined to notice randomness of the characters' positions and orientations when viewing a scene from this level. Therefore, less computation could be spent on characters in the background of scenes when the camera is at eye level and can be focused on the behavior of individuals in the foreground to increase realism. It remains to be seen, however, if this result holds for dynamic scenes.

Posthoc analysis using Newman-Keuls comparisons between means showed that viewpoint 1 (eye-level) was rated as real significantly more often than viewpoint 2 (canonical), but there were no differences between either viewpoint 3 (isometric) or viewpoint 4 (top-down). This suggests that, in general, using an eye-level view will result in a more realistic representation of pedestrian crowd formations to a viewer. Figure 19 shows participants' responses for each viewpoint across both open and corridor locations and random and context formations.

7. CONCLUSIONS & FUTURE WORK

In Experiment 1, participants judged open location scenes utilizing our contextual rules to be as real, or more real, than the real scenes. This could be explained by the fact that due to the large number of entrances and exits and the space available, a wider variety of formations were acceptable as being realistic, since fewer constraints could be applied by the viewer. In general, for the corridor location, for orientation only and position and orientation blocks, participants were better able to discern between the real and artificial, which could perhaps be due to the constrained nature of the zone. Participants were particularly sensitive to the position of characters in this location, as can be seen from the first experiment block, something that needs to be considered in more detail in future experiments. Despite this, overall, our combined contextual rules for this location were judged to be nearly as realistic as the real scenes. The fact that most of the participants were familiar with the areas used for both scenes, in particular the open location, implies that they were aware of the general directions of flows of pedestrians, and this is possibly reflected in their judgements of realism throughout the experiment.

In Experiment 2, we found that there was no difference in how participants perceived the realism of four different real scenes, implying that our results obtained in Experiment 1 were not due to any peculiarities in any one individual real image, but rather to the effects of all formation types. We also found that context scenes were judged to be equally as realistic or more realistic than real scenes. One possible explanation for this is that when people are familiar with an area, they may look for specific context elements of a static scene (e.g., out-of-bounds areas, obstacles) when making judgements on realism. These rules are obeyed by each character in our context scenes. However, in any real-life situation, while these context rules will be adhered to by the general population, there is an inherent element of randomness to human behavior, and a small number of people may be acting outside of these context rules (e.g., abrupt turning, facing in inappropriate directions). Our results suggest that if this is the case, that it was not a single image that conveyed this phenomenon, but that it was present across all of the real images viewed. In Experiment 2, we found a result that differed from Experiments 1 and 3; we found no effect of location on participants' perception of realism. We feel that this is due to the difference in the random placement rules used for this experiment. Placement of characters in out-of-bounds areas would have been more noticeable for the corridor location, since the ratio of out-of-bounds to in-bounds area was much higher than for the open location. Therefore, random placement of characters easier for participants to detect as synthetic scenes for corridor locations in Experiments 1 and 3.

In Experiment 3, where we examined the effects of camera viewpoint on the ability to discern between contextual and random scenes, we found that people are less able to identify randomness in the formation of characters when the camera is at eye-level. However, when viewed from a canonical viewpoint with more information available to the viewer, they were better able to better identify these random formations as being unrealistic. When the formations were created taking the context of the scene into account and the characters obeyed the natural social cues of the area, participants regarded these scenes as realistic, on average, 70% of the time, regardless of the camera viewpoint. These results are useful in two ways: We can deduce that our context rules are effective in both corridor and open locations no matter what angle they are viewed at, indicating that they are appropriate for initializing a crowd system; and when the camera is at eye-level, the accuracy of the positions and orientations of the crowd is not easily as perceptible to the viewer. This opens opportunities for reducing the simulation burden when certain behaviors of our agents do not need to be fully accurate and could possibly apply to other aspects such as rendering and animation of the characters also.

7.1 Recommendations

From the results we have obtained from our two perceptual studies, the following recommendations can be made for applying our results to crowd synthesis.

(1) Recommendation 1. The context of the location is an important consideration when creating pedestrian formations in a crowd scene.
(2) Recommendation 2. The context rules presented here are more important when the location is of a more constrained nature. When the scene is less constrained, less attention may be required for plausible positions and orientations due to the more open nature of the location and possibility for more behaviors.
(3) Recommendation 3. When the camera is at eye-level, less processing may be required to generate plausible positions and orientations than when the camera is at an angle that provides a more visible view of the individuals in the scene.

To summarize, our results show that a viewer's ability to distinguish between real and artificial scenes depends heavily on the context of the scene and how characters adhere to this context (e.g., characters walking bidirectionally in a constrained location). Our results indicate that contextual factors are vital when considering the perceived realism of pedestrian formations. We also show that the angle from which a scene is viewed has an effect on realism when these context cues are not taken into account. From a modeling standpoint, the results from these experiments, most significantly those taking both position and orientation into account, imply that the contextual rules presented here form an effective general starting point from which to populate urban environments. Applications of this work include creating initial formations of characters for populating real-time virtual environments and the placement of pedestrians in static urban architectural displays. There are also longer-term possibilities for application to level-of-detail metrics for pedestrian and crowd behavior models (see Shao and Terzopoulos [2005] and Pelechano et al. [2007]), where accuracy of simulation may be traded for performance in order to display larger crowds than would otherwise be possible while minimizing perceived errors in the behavior.

Nevertheless, many improvements and perceptual investigations remain to be made. Of particular importance is a more detailed investigation of factors at the group level, for example, group sizes, number and distribution of groups and group formations under both static and dynamic conditions. It seems likely that groups play an important role in the perception of pedestrian scenes. This is evident from Experiment 2, where our random placement of characters was only random in terms of the characters' orientation and the lack of grouping. Even when constraining random placement to in-bounds areas, like context rules, participants were able to identify random behaviors based on these two factors. Future work will examine these two factors to distinguish exactly how important groups are for realism in these scenes. No doubt there is a vast array of factors for possible consideration when creating context rules for groups: It will be challenging to identify, isolate and choose a subset of these in order to be able to conduct tractable experiments.

We also plan to consider dynamic pedestrian scenarios and study the formulation of dynamic context rules derived from those for static scenes, as described in this article. Perceptual evaluation of such rules and comparisons with existing models, for example, those based on steering behaviors [Reynolds 1987], could be of great utility in the evaluation and construction of crowd simulations for interactive applications.

REFERENCES

BLANZ, V., TARR, M. J., BULTHOFF, H. H., AND VETTER, T. 1999. What object attributes determine canonical views? *Perception 28*, 575–600.

BULTHOFF, H. H., EDELMAN, S. Y., AND TARR, M. J. 1995. How are three-dimensional objects represented in the brain? *Cerebral Cortex 5*, 3, 247–260.

CHENNEY, S. 2004. Flow tiles. In *Proceedings of the Eurographics Symposium on Computer Animation (SCA'04)*. ACM, New York, 233–242.

ENNIS, C., PETERS, C., AND O'SULLIVAN, C. 2008. Perceptual evaluation of position and orientation context rules for pedestrian formations. In *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization (APGV'08)*. ACM, New York, 75–82.

GALLESE, V., FADIGA, L., FOGASSI, L., AND RIZZOLATTI, G. 1996. Action recognition in the premotor cortex. *Brain 119*, 593–609.

GOMEZ, P. AND SHUTTER, J. 2008. Memory for objects in canonical and non canonical viewpoints. *Psychonomic Bull. Rev. 15*, 5, 990–944.

HELBING, D. AND MOLNÁR, P. 1995. Social force model for pedestrian dynamics. *Phys. Rev. E 51*, 5, 4282–4286.

HENRY, D. AND FURNESS, T. 1993. Spatial perception in virtual environments: Evaluating an architectural application. In *Proceedings of the Annual International Virtual Reality Symposium*. IEEE, Los Alamitos, CA, 33–40.

LAMARCHE, F. AND DONIKIAN, S. 2004. Crowd of virtual humans: a new approach for real time navigation in complex and structured environments. *Comput. Graphics Forum 51*, 3, 509–518.

LEE, K. H., CHOI, M. G., HONG, Q., AND LEE, J. 2007. Group behavior from video: A data-driven approach to crowd simulation. In *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA'07)*. ACM, New York, 109–118.

LERNER, A., CHRYSANTHOU, Y., AND DANI, L. 2007. Crowds by example. *Comput. Graphics Forum 26*, 3, 655–664.

LYONS, M. J., CAMPBELL, R., PLANTE, A., COLEMAN, M., KAMACHI, M., AND AKAMATSU, S. 2000. The noh mask effect: vertical viewpoint dependence of facial expression perception. *Proc. Biol. Sci. 267*, 1459, 2239–2245.

MCDONNELL, R., NEWELL, F., AND O'SULLIVAN, C. 2007. Smooth movers: Perceptually guided human motion simulation. In *Proceedings of the Eurographics Symposium on Computer Animation (SCA'07)*. ACM, New York, 259–269.

MUSSE, S. R. AND THALMANN, D. 1997. A model of human crowd behavior: Group interrelationship and collision detection analysis. In *Proceedings of the Eurographics Workshop on Animation and Simulation.* Springer-Verlag, Berlin, 39–52.

PALMER, S. E. AND ROSCH, P. C. 1981. Canonical perspective and the perception of objects. In *Proceedings of the 9th International Symposium on Attention and Performance*. 135–151.

PELECHANO, N., ALLBECK, J. M., AND BADLER, N. I. 2007. Controlling individual agents in high-density crowd simulation. In *Proceedings of the Eurographics Symposium on Computer animation (SCA'07)*. ACM, New York, 99–108.

PETERS, C., ENNIS, C., MCDONNELL, R., AND O'SULLIVAN, C. 2008. Crowds in context: Evaluating the perceptual plausibility of pedestrian orientations. In *Proceedings of Eurographics Short Papers*. Springer-Verlag, Berlin.

PREMACK, D. AND WOODRUFF, G. 1978. Does the chimpanzee have a theory of mind? *Behav. Brain Sci. 1*, 515–526.

REITSMA, P. S. A. AND POLLARD, N. S. 2003. Perceptual metrics for character animation: sensitivity to errors in ballistic motion. *ACM Trans. Graph. 22*, 3, 537–542.

REYNOLDS, C. W. 1987. Flocks, herds, and schools: A distributed behavioral model. *Comput. Graphics 21*, 4, 25–34.

SCHILBACH, L., WOHLSCHLAEGER, A., KRAEMER, N., NEWEN, A., SHAH, N., FINK, G., AND VOGELEY, K. 2006. Being with virtual others: Neural correlates of social interaction. *Neuropsychologia 44*, 5, 718–730.

SHAO, W. AND TERZOPOULOS, D. 2005. Autonomous pedestrians. In *Proceedings of the Eurographics Symposium on Computer Animation (SCA'05)*. ACM, New York, 19–28.

TARR, M. J. 1995. Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bull. Rev. 2*, 1, 5582.

YARBUS, A. L. AND HAIGH, B. 1973. *Eye Movements and Vision*. Plenum Press, New York.

Received February 2009; revised February 2010; accepted March 2010