PedVR: Simulating Gaze-Based Interactions between a Real User and Virtual Crowds

Sahil Narang*
Universty of North Carolina Chapel Hill
Andrew Best†
University of North Carolina Chapel Hill
Tanmay Randhavane‡
University of North Carolina Chapel Hill
Ari Shapiro§
USC Institute for Creative Technologies
Dinesh Manocha¶
University of North Carolina Chapel Hill

http://gamma.cs.unc.edu/pedvr

Figure 1: *Our algorithm generates plausible full body motion for tens of virtual agents and allows the user to interact with the virtual crowd. (Left) The user is provided with a first person view through an HMD. (Center) The virtual characters display plausible behaviors such as gazing and gesturing. (Right) The real user (shown in blue) can freely move in the virtual world while the agents actively avoid collisions. We highlight the gaze using line-of-sight between real user and a virtual user.*

Abstract

We present a novel interactive approach, *PedVR*, to generate plausible behaviors for a large number of virtual humans, and to enable natural interaction between the real user and virtual agents. Our formulation is based on a coupled approach that combines a 2D multi-agent navigation algorithm with 3D human motion synthesis. The coupling can result in plausible movement of virtual agents and can generate gazing behaviors, which can considerably increase the believability. We have integrated our formulation with the DK-2 HMD and demonstrate the benefits of our crowd simulation algorithm over prior decoupled approaches. Our user evaluation suggests that the combination of coupled methods and gazing behavior can considerably increase the behavioral plausibility.

Keywords: multi-agent simulation, crowds, human agents, virtual reality

Concepts: •**Human-centered computing** → **Interaction paradigms;** •**Computing methodologies** → **Computer graphics;** *Animation;*

1 Introduction

Virtual Reality (VR) is increasingly being used for a wide range of applications including computer-aided design, architectural and urban walkthroughs, entertainment, virtual tourism, telepresence, etc. There has been considerable progress toward increasing the sense of realism in virtual worlds in terms of scene complexity, visual rendering, acoustic effects, physics-based simulation, and interaction paradigms. However, current virtual worlds tend to be mostly static and one of the major challenges is to simulate plausible virtual humans or crowds. It is known that the presence of human-like agents can improve the sense of immersion [Pelechano et al. 2008; Llobera et al. 2010; Slater et al. 2006]. This is important for training simulators [Ulicny and Thalmann 2001; Romano and Brna 2001], architectural flow analysis and evacuation planning, virtual reality therapy for crowd phobias, social anxiety [Pertaub et al. 2002] and PTSD treatments [Rothbaum et al. 2001].

Many of these applications need capabilities for a real user to be immersed as part of the virtual crowd. This includes simulating the movements and behaviors of large numbers of virtual agents at interactive rates and developing natural interaction mechanisms between the user and virtual agents. The problem of simulating virtual humans and crowds has been extensively studied in computer animation, virtual reality, robotics, and pedestrian dynamics. Many methods have been proposed for computing collision-free trajectories for 2D agents in a plane, for human motion synthesis, and for real-time rendering of crowds on commodity hardware. However, it is quite challenging to generate plausible simulations of a large

---

*sahil@cs.unc.edu
†best@cs.unc.edu
‡tanmay@cs.unc.edu
§shapiro@ict.usc.edu
¶dm@cs.unc.edu

VRST '16,, November 02-04, 2016, Garching bei München, Germany
ISBN: 978-1-4503-4491-3/16/11
DOI: http://dx.doi.org/10.1145/2993369.2993378

group of human-like agents, especially in dense scenarios with obstacles. Each agent is typically modeled using tens of degrees-of-freedom (DOF). The resulting high-dimensional motion trajectories need to satisfy various constraints, such as collision-free, biomechanics constraints, stability, natural-looking motion, etc. In addition, we need capabilities for a real user to walk among the virtual agents in a natural manner and avoid collisions. Finally, we need the ability to communicate with virtual agents using different cues such as gaze or eye contact [Bailenson et al. 2005].

Current approaches for interactive crowd simulation are based on computing collision-free trajectories for 2D agents (e.g. circles) in a plane. The resulting full-body agent motions are computed by synthesizing 3D motions for human-like characters that follow the 2D trajectories. However, these approaches have many shortcomings and may not result in plausible simulations in many scenarios. The simplified 2D navigation methods cannot possibly account for the entire range of human motion and kinematic or bio-mechanical constraints. The resulting combination with a high DOF motion synthesis system can lead to artifacts such as foot skating, bone stretching, unnatural balance, etc. Furthermore, these decoupled trajectory computation and motion synthesis approaches cannot account for many interactions between the real user and virtual agents.

In addition to trajectory computation and collision-free interactions, it is important to exhibit human like behaviors such as gazing and gesturing. Gaze, in particular, is a key aspect of non-verbal communication [Bailenson et al. 2005]. Recent studies have indicated the effect of gaze in terms of interpretation of emotional expressions [Adams and Kleck 2003; Gallup et al. 2014]. Gaze is also increasingly being used by embodied conversational agents (ECA) to increase the believability and thereby, the plausibility of the simulation [Peters et al. 2005]. However, current interactive crowd simulation methods are unable to simulate such non-locomotive or communication behaviors.

Main Results: We present a novel interactive approach, *PedVR*, to generate plausible behaviors of a large number of virtual humans and to facilitate natural interactions between the real user and virtual agents. Our formulation is based on a coupled high-dimensional trajectory computation algorithm that combines 2D navigation methods with an interactive human motion synthesis algorithm [Narang et al. 2016]. The resulting approach can yield more human-like trajectories and collision-free navigation between the virtual agents. Furthermore, we account for the presence of a tracked real user in the shared virtual environment and generate plausible trajectories in an asymmetric manner. In addition, we present novel techniques to generate plausible upper body motion for each virtual agent that supports gazing and gesturing, and also increases the behavioral realism of virtual characters. Different behaviors are specified and triggered using a Behavioral Finite State Machine (BFSM). To generate interactive simulations, we parallelize many stages of our algorithm on multiple cores. We demonstrate the performance of our system on several scenarios with tens of virtual agents.

We have evaluated the level of presence achieved by a real user immersed in an environment composed of virtual humans. In particular, we compared the following algorithms to showcase the benefits of our coupled high-dimensional agent trajectory computation algorithm:

- **Decoupled** : A widely used decoupled 2D navigation algorithm [van den Berg et al. 2011].
- **PedVR** : Our novel coupled high dimensional trajectory computation algorithm.
- **PedVR+G** : Our coupled high dimensional trajectory computation algorithm with the addition of gazing behavior.

We conducted a within-subjects user study with 20 subjects and performed the evaluations in two scenarios using the DK-2 head mounted display. Our studies to measure the level of presence are based on prior work on evaluating crowd simulation algorithms [Pelechano et al. 2008; Garau et al. 2005]. Our results indicate that subjects prefer PedVR to Decoupled in 41.3% of responses with 8.8% of responses indicating strong preference and 31.9% indicating no difference between the two. With the introduction of gaze behaviors, we see a preference for PedVR+G in 56.2% of responses with 35.6% indicating a strong preference and only 10% indicating no difference. Our results indicate a 4 fold increase in the number of strong preferences when gaze behaviors are presented. In all cases we see a statistically significant preference for PedVR.

We also demonstrate the capabilities of our system on a number of complex indoor and outdoor real-world-like environments.

The rest of the paper is organized as follows. In Section 2, we survey related work in crowd simulation and motion synthesis for human-like agents. We present an overview of the approach and the details of our coupled planning and motion synthesis algorithm in Section 3. We present the details of the user interaction, including collision-free motion and gaze, in Section 4. We provide implementation details and highlight the performance of our framework on several benchmarks in Section 5. We describe our evaluation framework and discuss the relative benefits of coupled full-body trajectory computation and gaze in Section 6.

2 Related Work

In this section, we give a brief overview of prior work on multi-agent simulation, motion synthesis and crowd simulation for VR.

2.1 Multi-Agent Crowd Simulation

Most prior 2D crowd simulation techniques can be broadly classified as macroscopic models and microscopic models. Macroscopic models [Treuille et al. 2006] tend to compute the aggregate motion of the crowd by generating fields based on continuum theories of flows. Microscopic models based on multi-agent methods compute trajectories for each individual agent. These use a combination of global [Snook 2000] and local navigation methods [Helbing et al. 2000; Karamouzas et al. 2014; van den Berg et al. 2011; Schadschneider 2002] then adapts the plan to avoid collisions with other agents and dynamic obstacles. Most of these methods only compute the trajectories of the agents in a 2D plane.

2.2 Human-like Motion Synthesis

There is extensive literature in computer graphics and animation on generating human-like motion [Welbergen et al. 2010]. We limit our discussion to data-driven, procedural, and physics-based methods. Data-drive methods such as motion graphs [Kovar et al. 2002; Feng et al. 2012] create a parameterized graph of blendable motions and apply traversal algorithms to generate trajectories. Such motion databases are often created through motion capture yielding human-like results. Procedural methods apply kinematic principles to generate motions adhering to bio-mechanic constraints [Bruderlin and Calvert 1993]. Physics-based models seek to generate physically-feasible motions by computing actuator forces for each joint to generate the desired motion [Jain et al. 2009]. These methods generate physically correct motions but may not generate natural motions.

2.3 Multi-agent Simulation & Motion Synthesis

There are few methods that combine crowd simulation and motion synthesis into one framework. Shapiro et al. [2011] present a character animation framework that utilizes a 2D steering algorithm and a motion blending-based technique to generate visually appealing motion. ADAPT [Kapadia et al. 2014] combines an open-source navigation mesh and steering algorithm with a set of animation controllers. There is work in the robotics domain that addresses bi-pedal locomotion for multiple robots [Park et al. 2015], though they are not fast enough for interactive applications.

2.4 Crowd Simulation for VR

There is little work in simulating crowds in VR applications. Pelechano et al. [2008] performed user evaluations where the subjects were free to move around in a virtual environment that was populated with agents; and the trajectories of the agents were computed using different algorithms. They used presence questionnaires to evaluate different crowd models. Llobera et. al. [2010] measured electrodermal responses of subjects as they were approached by virtual characters in VR. In their set, the user or subject was static in the virtual scene and the experimental setup prevented any collisions. Kiefer et al. [2013] discussed tradeoffs between different VR methodologies with respect to mobility rehabilitation. Cirio et al. [2013] compared various interfaces for locomotion in VR by comparing the virtual trajectories to real trajectories. Kim et al. [2016] presented a data-driven method that used trajectories extracted from videos to simulate the motion of virtual agents. They evaluated the benefits of their approach by comparing it with synthetic multi-agent models. Bonsch et al. [2016] studied the effects of variations in gaze and avoidance maneuvers of a single virtual agent in a small office setting. There is considerable work on embodied conversational agents (ECA) [Von Der Pütten et al. 2009; Cassell 2001], in which animated anthropomorphic interface agents are used to engage a user in real-time, multimodal dialogue, using speech, gesture, gaze, posture, and other verbal and non-verbal behaviors. In most cases, ECA is restricted to one user-agent interaction. Other methods have attempted to insert virtual crowds as an overlay on a real-world video [Rivalcoba et al. 2014; Ren et al. 2013].

3 Interactive Crowd Simulation

In this section, we introduce the notation and terminology used in the rest of the paper. We also give an overview of our coupled approach, PedVR, that combines 2D multi-agent navigation and 3D human motion synthesis, and can generate gazing behaviors.

3.1 Notation and Assumptions

Let $\mathbf{S}$ represent the simulator state, defined as the union of all entities in the scene, including obstacles in the environment and the overall state space $\mathbf{Q} = \cup_i \mathbf{q}_i.$, where $\mathbf{q}_i$ denotes the state space of each agent i. An agent i in our simulation has an associated skeletal mesh that is used for high-DOF trajectory computation. Each configuration $\mathbf{q}_i$ of the skeletal mesh is defined using the degrees-of-freedom (DOF) used to specify the 6-DOF root pose and the joint angles in a n-dimensional vector space. The trajectory in this high dimensional configuration space is a function of time, and is denoted as $\mathbf{q}_i(t)$.

We project the geometric representation of each skeletal mesh in $\mathbb{R}^n$ space to $\mathbb{R}^2$ space and bound it with a tightly fitted circle of radius r_i. This circle is used by the 2D multi-agent navigation algorithm. Thus, each skeletal mesh, with a 6-DOF root joint denoted $\mathbf{q}_i^j$, is represented in the 2D simulator by a circle of radius r_i, positioned at $\vec{p}_i$, where $\vec{p}_i$ is simply the projection of the root joint, $\mathbf{q}_i^j$, on the 2D XY plane. The 2D navigation algorithm generates trajectories that correspond to the XY-root translation of the 6-DOF root joint $\mathbf{q}_i^j$ of the associated skeleton. These collision-free trajectories are denoted as 2D time-varying functions representing position $\vec{p}_i^c(t)$ and velocity $\vec{v}_i^c(t)$. At any given instant, these functions can be sampled to yield the 2D collision-free position $\vec{p}_i^c$ and velocity $\vec{v}_i^c$ of the corresponding disc agent. The user's input is mapped to a "user agent" and is assymmetrically avoided by virtual agents.

Figure 2 provides an overview of our approach and shows how various components relate to behavior specification, 2D navigation, 3D human motion synthesis, gaze generation, and the integration with the immersive hardware and the game engine used for rendering. Our coupling approach uses a multi-level 2D navigation algorithm integrated with a 3D human motion synthesis module for high-DOF articulated bodies based on a closed feedback loop. Such an approach allows us to simulate tens of virtual agents at interactive rates on current multi-core CPUs, and also to generate plausible behaviors in terms of collision-free trajectories, natural passing of agents, and gaze computation.

3.2 2D Multi-Agent Simulation

As described above, agents are modeled as two-dimensional disks of radius r. We use a multi-agent approach, i.e., each agent is modeled as a discrete entity, capable of planning and acting on its own. We use a Behavioral Finite State Machine (BFSM) that maps the current simulation state and time to a goal position $\vec{g}_i$ for each agent i in the simulation. Given the current goal position of an agent, we decompose the 2D trajectory computation problem into two phases: global path planning and local navigation.

3.2.1 Global Path Planning

The global planner can be represented by the function $\mathbf{P}_i : \mathbb{S} \times \mathbb{R}^2 \rightarrow \mathbb{R}^2 \times \mathbb{R}^2$, which maps the simulator state and the agent's goal position into an instantaneous preferred velocity, $\vec{v}_i^o$, and preferred orientation, o_i^o of that agent. This velocity and orientation are used to specify the movement of the agent. The global planner is used to compute the collision-free path to the goal with respect to static obstacles in the simulation. This path is communicated to the local planner as the "preferred velocity", $\vec{v}^o$, and "preferred orientation", o^o. We use a precomputed navigation mesh [Snook 2000] that decomposes the traversable space into connected, complex polygons and generates intermediate way points. Our formulation makes the assumption that each agent is always facing its intermediate way-point as long as the way-point is visible. We use a kd-tree to perform visibility queries and set the preferred orientation of the agent to face toward the visible way-point.

3.2.2 Local Navigation

Let $\mathbf{LCA}_i : \mathbb{S} \times \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}^2$ denote a local collision avoidance function that maps the simulator state, the instantaneous preferred velocity, and time horizon, τ, into a collision-free 2D velocity, $\vec{v}_i^c$, with respect to other agents in the environment for at least time τ. In other words, it tends to compute a velocity that can generate a collision free trajectory for time τ. We utilize an efficient 2D collision avoidance model that can generate smooth, stable, collision-free velocities efficiently and is thus, ideally suited for VR applications. We provide details of the model in Section 4.

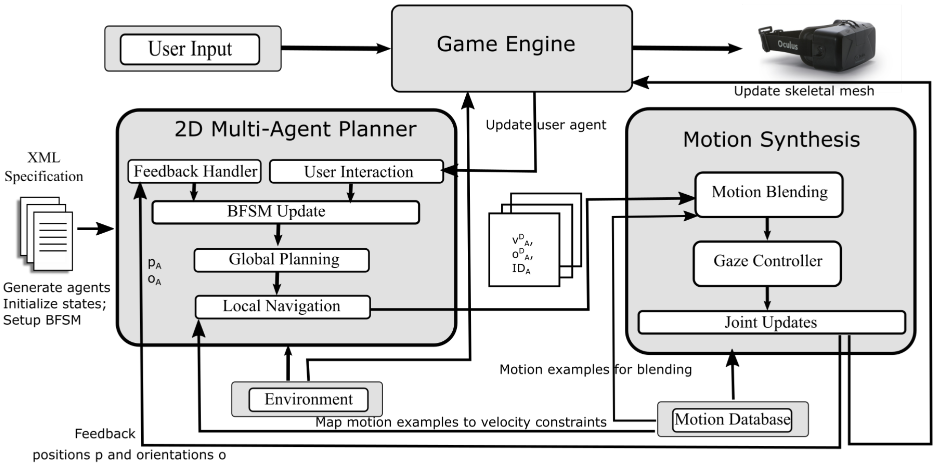

Figure 2: ***System Overview.*** *We use a coupled approach to generate full body motion for multiple agents. The user's input is mapped to a "user agent" at every timestep. The 2D planner leverages a human motion database and generates collision-free velocities while asymmetrically avoiding the user. The motion synthesis module generates appropriate upper and lower body motion. Finally, the user is presented with a first person view of the virtual world with an HMD.*

3.3 Motion Synthesis

The motion synthesis module is responsible for computing the trajectory $\mathbf{q}_i$ for the articulated skeleton in *n*-dimensional spaces. We utilize the character animation package, Smartbody [Shapiro 2011], to generate plausible locomotive and non locomotive motion. Further details are provided in Section 4.

3.4 Coupled 2D Navigation & 3D Motion Synthesis

The low dimensionality of the 2D planning space implies that the 2D collision-free velocity $\vec{v}_i^c$ may not satisfy different human motion constraints, including kinematic constraints and bio-mechanical constraints. Therefore, the resulting high-dimensional trajectory $\mathbf{q}_i$ of the articulated skeleton is likely to introduce some variability in the synthesized velocity of the root joint $\mathbf{q}_i^j$ and this may lead to collisions or other artifacts. We overcome this issue by incorporating human motion constraints in the 2D multi-agent navigation algorithm (Section 4). Moreover, we synchronize the 2D agent positions with their corresponding articulated skeletons at the beginning of each simulation step.

3.5 User Interaction

Our framework is agnostic to the specific input method used to track the user's movement in the virtual environment. The user is free to move around in the virtual environment populated with virtual agents. The user could be walking or using a keyboard or joystick for navigation. The user's input is mapped to a special "user agent", denoted by q_u. More details on virtual agent-user interactions are provided in Section 4.

3.6 Simulation Update

At the beginning of every simulation step, the 2D disc agents are updated to reflect the positions and orientations of their corresponding skeleton. In case of each user agent, we synchronize the velocity in addition to the position and orientation. The 2D navigation algorithm leverages the motion database of precomputed or recorded human motions, and computes a collision-free velocity, $\vec{v}_i^c$, orientation, $\vec{o}_i^c$, and a BFSM state, ID_i, for each virtual agent *i*. This information is communicated to the motion synthesis module that updates the skeleton. In addition, the user's input is used to update the corresponding position of the user agent's skeleton. The skeletal information for each agent is transferred to the rendering engine. Finally, the user is provided the view from the camera that is positioned at the base of the neck of its corresponding skeleton through an HMD.

4 Interactions between a Real User and Virtual Agents

In this section, we present details on our virtual human agent simulation algorithm. Moreover, we present various techniques that can improve the interaction between virtual agents, and between the real user and virtual agents in an immersive environment. It is imperative that the virtual agents exhibit plausible human-like behavior to enhance the believability of the virtual world and prevent breaks in presence [Slater and Steed 2000; Slater et al. 2006]. First, the virtual agents must navigate in the environment and avoiding collisions with the user, other virtual agents, and the obstacles in the scene. Second, the user should be able to interact with nearby virtual agents and communicate in an explicit or implicit manner.

4.1 Collision-Free Navigation & Motion Synthesis

Current approaches for crowd simulation tend to generate 3D motion for human-like characters as a post-processing step. These motions follow the 2D trajectory rigidly and may result in awkward or implausible motions such as foot skating, bone stretching, unnatural balance, etc. The simplified 2D navigation methods cannot possibly account for the entire range of human motion and kinematic or bio-mechanical constraints.Furthermore, these decoupled trajectory computation and motion synthesis approaches cannot account for many interactions between the real user and virtual agents that require dynamic planning and motion synthesis.

We utilize a coupled motion synthesis approach with combines a “social-force” based method [Karamouzas et al. 2014] and reciprocal velocity obstacles [Van den Berg et al. 2008]. In addition, we introduce constraints based on the dynamic-constraints of the skeletal mesh. Our algorithm generates 2D trajectories which guarantee collision avoidance and generate motions feasible for articulated agents. The velocity and orientation computed by the navigation algorithm are used to synthesize appropriate human motion using a motion blending technique [Feng et al. 2012]. A thorough explanation and analysis our navigation algorithm is provided in [Narang et al. 2016].

4.2 Gazing

Gaze is an important aspect of human face to face interaction, and can be used to increase the behavioral plausibility of the virtual characters and the overall quality of the immersive virtual experience. We begin by determining if the user agent, u, is visible w.r.t virtual agent i. For the sake of computational efficiency, we do not consider partial visibility, restricting the visibility query to two dimensional space.

We then determine if the user agent is heading towards the virtual agent, using the following set of equations:

$$\hat{d} = \frac{\vec{p}_u - \vec{p}_i}{\|\vec{p}_u - \vec{p}_i\|}, \tag{1}$$

$$w = \vec{v}_i.\hat{d}. \tag{2}$$

Let g denote a boolean that denotes whether agent i should gaze at user agent u, given by:

$$g := (\|\vec{p}_u - \vec{p}_i\| < \mathbf{D}_1) \wedge (w > 0) \wedge (w < \mathbf{D}_2) \wedge v_{ui}, \tag{3}$$

where $\mathbf{D}_1$,$\mathbf{D}_2$ are pre-defined constants representing a maximal gaze distance and approach speed envelope respectively, and v_{ui} denotes the visibility of agent u with respect to agent i.

In cases for which g evaluates to true, we use the gaze controller present in Smartbody [Thiebaux et al. 2009] which is capable of producing gaze shifts with configurable styles. It does so by manipulating a set of joints of the skeletal mesh, subject to kinematic and smoothing constraints. The gaze computation is performed by each virtual agent w.r.t to only the user agent and gaze is maintained as long as these conditions remain true.

4.3 Gestures & Upper Body Motion

In addition to gaze, a virtual character may exhibit gestures and non locomotive behaviors. These gestures may be triggered using a Behavioral Finite State Machine (BFSM). We use a BFSM to represent the mental state (including such factors as immediate goals and mood) of agents in the simulation. The BFSM can be represented by a function $\mathbf{B}_i : t \times \mathbb{S} \rightarrow \mathbb{I} \times \mathbb{R}^2$ which maps the time and simulator state into a unique BFSM state, ID, and corresponding goal position $\vec{g}_i$ for agent i. Furthermore, we define a mapping $\mathbf{G} : \mathbb{ID} \rightarrow \mathcal{M}$ where the set $\mathcal{M}$ denotes a set of gestural motions $m^k(i = 1...k)$. During the simulation, an arbitrary gesture selection policy may be applied to select a motion, $\mathbf{m}^k \in \mathcal{M}_{ID}$. Thus, we can simulate diverse and complex behaviors. For example, the BFSM is used in the tradeshow scene (Section 5) to select a goal ”booth” based on a probabilistic distribution. Once the agent arrives at its goal booth, it waits 10-20 seconds before choosing another booth. Furthermore, the agent may turn and gaze at the user if the user is ”too close” to the agent. Such complex behaviors can be easily implemented using the BFSM.

Benchmark	Agents	Average frame update time (ms)		
		Decoupled	PedVR	PedVR+G
Shopping Mall	24	15	18	25
Shibuya	32	16	18	22
Tradeshow	30	18	22	24
Anti-podal Circle	8	8	16	10
Bidirectional Flow	8	16	15	20

Table 1: ***Average Frame Update:*** *Each virtual agent has 38 joints. Our framework can simulate 30+ agents at 40-60 fps. Timing results were gathered on a Intel Xeon E5-1620 v3 with 4 cores and 16 GB of memory.*

5 Performance Evaluation

We have implemented our algorithm in C++ on a Windows 7 desktop PC. All the timing results in the paper were generated on an Intel Xeon E5-1620 v3 with 4 cores and 16 GB of memory. We demonstrate the performance of our algorithm on the following benchmark scenarios and provide running times in Table 1:

Shibuya Crossing We simulate a busy street crossing (Figure 3(a)), where each agent is initialized at a different position of the intersection. The BFSM is used to assign distinct goal positions for each agent based on a probabilistic distribution. Agents reach their goals, wait for a few seconds, and then move towards another goal. In most cases, PedVR agents exhibit smooth collision avoidance behaviors while avoiding the user agent. However, overt collision avoidance behaviors, such as sidestepping and turning, can be observed if the user agent suddenly or aggressively approaches the virtual agent (Figure 4(a)(b)). Our system can simulate 30+ agents at approx. $45 - 60$ fps.

Tradeshow This is a challenging scenario for any crowd simulation algorithm. It highlights the environment corresponding to a tradeshow (Figure 3(b)) with several obstacles and narrow passages. Agents walk up to randomly assigned booths, spend a few seconds there, and then move to another booth. Agents can be seen smoothly avoiding collisions with one another in the narrow passages. Despite the large number of obstacles and narrow passage constraints, our system can simulate 30 agents at $40 - 50$ fps.

Shopping Mall This scenario shows a shopping mall where agents walk around the shops and pass each other in the narrow hallways (Figure 3(c)) similar to the tradeshow scenario. agents may stop at some shops. Overall, we observe smooth trajectories and collision avoidance behaviors. Our system can simulate tens of agents at $40 - 55$ fps.

6 User Evaluation

In this section, we detail a within-subjects user study conducted to evaluate our method (PedVR) with and without gaze, compared to a baseline decoupled crowd simulation algorithm.

6.1 Experiment Goals and Expectations

Our experiment sought to determine whether or not significant benefits could be attained by using our coupled algorithm as compared to a decoupled method. We expected to find that participants consistently indicate preference for our algorithm over a baseline and that the presence or absence of gaze behaviors would yield substantive changes to the level of preference for the method.

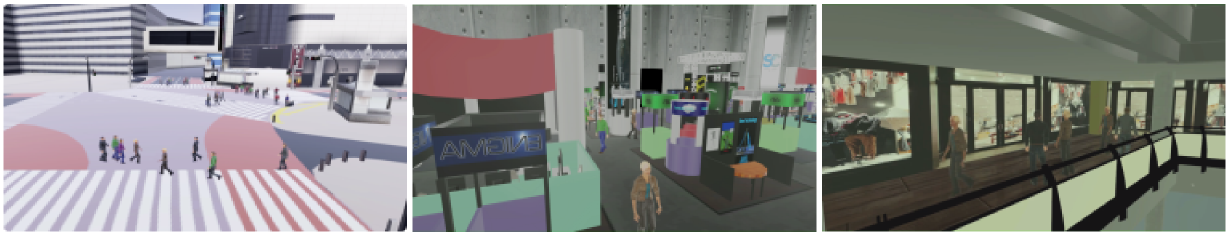
(a) Shibuya Crossing (b) Tradeshow (c) Shopping Mall

Figure 3: Benchmarks: *We simulate several real world complex scenarios with 30+ agents at interactive rates. Our algorithm generates plausible full body motion for multiple virtual agents using a coupled planning and motion synthesis approach.*


A B C D

Figure 4: Virtual Agent-User Interactions: *PedVR agents can take overt collision avoidance measures, such as (A) sidestepping and (B) turning to avoid sudden or aggressive movement by the user. For visual clarity, the user agent is visualized in blue from overhead. (C) Agents gaze at the user as they pass by and (D) are also capable of gesturing.*

6.2 Experimental Design

The study was designed as a within-subjects study in which participants would experience each of the three evaluated methods using an oculus DK-2 head-mounted display and a mouse and keyboard for virtual movement. In each of the scenarios outlined below, participants were tasked with following a red-sphere through a virtual world populated with virtual agents. A following task was chosen to reduce variability in the amount of time users spent exploring the space independent of which simulation method was being evaluated. The participant was presented with three trials for each scenario, corresponding to method with which the virtual agents were simulated. Our study was conducted in person in a laboratory setting.

6.2.1 Evaluated Methods

In the study, participants compared three different full body simulation algorithms:

- **PedVR : PedVR without Gaze** We use a coupled approach for 2D navigaiton and fully body motion synthesis, as described in Sections 3 & 4.
- **PedVR+G : PedVR with Gaze** We augment PedVR with gaze behavior, as described in Section 4.2.
- (**Decoupled**): We use a widely used decoupled method [van den Berg et al. 2011] for 2D navigation and motion blending for locomotive motion. As with the coupled approach, we sync the 2D agent with the root joint of the corresponding skeletal mesh for a fair comparison.

6.2.2 Scenarios

The users were presented with a total of three scenarios. Two scenarios were used for direct evaluation in this study and one was used for trajectory generation as part of a larger research effort. Figure 5 illustrates the scenarios and task. The scenarios were:

- **Antipodal Circle:** In this scenario, 10 virtual agents move to randomly sampled positions on the perimeter of a circle of radius 6 meters. The probabilistic goal assignment was designed to increase the density of agents at the center of the circle. The red sphere traveled between several points along the circumference of the sphere, to keep the user occupied for 60-90 seconds. Participants experienced this scenario using each of the three simulation algorithms described above.
- **Bidirectional Flow:** In this scenario, 8 virtual agents, in two groups of 4, moved towards each other from opposite ends of a 14 meter corridor. At each end, the groups turned around and crossed again. The red sphere was placed at the end of the corridor and moved to the other end as the participant reached towards its position. This scenario highlights the interactions during head-on collision avoidance behaviors and demonstrates the potential benefit of gaze behaviors. Participants experienced this scenario using each of the three simulation algorithms described above.
- **Head-on Corridor:** In this scenario, a virtual agent moves from one end of a narrow corridor towards the user with the red sphere position directly behind the virtual agent. Thus, the user was encouraged to walk head on towards the virtual agent. Participants experienced this scenario four times, each with a differing parameter for the PedVR algorithm, described in Section 6.4.

6.2.3 Variables

Independent: In this study, there are two independent variables. First, the scenario which the user is evaluating, and the second corresponds to the specific comparison they are making between the three evaluated methods.

Dependent: The dependent variable in the study is the participant's response to the social presence questionnaire (below) for each comparison in each scenario.

6.2.4 Metrics

There have been several approaches to measuring presence including self-reported questionnaires, behavioral responses, physiological responses [Slater et al. 2006] and breaks-in-presence (BIPs) [Slater and Steed 2000]. Physiological responses and BIPs may be more reliable than questionnaires, but are largely restricted to simulations with abrupt changes to induce such responses. Hence, for our study, we chose to utilize well established questionnaires.

Social Presence: Our evaluation primary relied on a modified version of the questionnaire introduced by Garau et al. [2005]. In our modification, a subset of the original questions were used and participants were not asked to directly rate the algorithms as in the original. Rather, for each question the participant indicated which method (if any) better represented the question in pairwise fashion. The methods were labeled in order of appearance for the user as A, B, and C respectively. Participants noted their preference using a 7 point Likert scale with values labeled (”Left much better”, ”Left Better”, ”Left Slightly Better”, ”No Difference”, ”Right Slightly Better”, ”Right Better”, ”Right Much Better”). In this response format, a value of one indicates a strong preference for the method listed on the left of the comparison. Table 2 gives the details about our specific questionnaire.

Simulator Sickness Index: As is common practice, we administered a simulator sickness questionnaire (SSQ) [Kennedy et al. 1993], before and after the study.

6.3 Participants

Participants were recruited on a university campus and consisted of graduate students and staff members. 20 participants were recruited: 7 females and 13 males, $M_{age} = 25$ years, $SD_{age} = 7.26$ years. Before agreeing to participate, they were given a high level overview of the setup and it was ensured that they felt comfortable using an HMD. The average time for conducting the study was about 35 minutes per participant. Participants were paid an equivalent of $10 for participation.

6.4 Procedure

Participants were welcomed and were instructed on the number of scenarios and the number of trials for each scenario. They signed a consent form and provided optional demographic information about their age and gender. Participants were then asked to fill the the Simulator Sickness questionnaire (SSQ).

After the SSQ, participants were presented with training scenario with no virtual agents, just the primary task of following the sphere. This was done to familiarize them with the HMD and the virtual controls. Participants then experienced the Antipodal Circle and Bidirectional Flow scenario in a counterbalanced order. In each scenario, participants experienced each of the target methods, also in a counterbalanced order. At the end of the third trial for each scenario, the participants were administered the social presence questionnaire detailed above.

Each participant then experienced the Head-on Corridor scenario four times, each corresponding to the virtual agent taking 0%, 25%, 50% and 100%, of the responsibility for avoiding collisions with the participant in random order. These trajectories will provide insight into whether the user perceived a need to avoid the virtual agent as they would other people in a typical narrow passage. Trajectories for this scenario were recorded. The participants were then administered the Simulator Sickness Index and allowed to provide feedback through a questionnaire and verbally with the experimenter.

In which simulation, did you have a greater sense of being in the same space as the characters?
In which simulation, did you respond to them as if they were real people?
In which simulation, did you make a greater effort to avoid the characters?
In which simulation, did the presence of the characters affect you more in the way you explored the space?
In which simulation, did the characters seem to respond to you more?
In which simulation, did the characters seem to look at you more?
In which simulation, did the characters seem to be more aware of you?
In which simulation, did you feel more observed by the characters?

Table 2: ***Questionnaire.*** *Questions presented to participants after each scenario [Garau et al. 2005]. Participants were asked to compare the three methods in pairs.*

6.5 Results and Discussion

In this section, we limit our analysis presented to studying the participant responses to the virtual scenarios and the three simulation algorithms described above. For each comparison between different pairs (PedVR+G / Decoupled, PedVR / Decoupled, PedVR+G / PedVR), we combined responses to the questionnaire into a single overall social presence preference index by computing the mean participant response to each comparison. We validated our aggregation by computing Cronbach's α for each questionnaire ($.78 < \alpha < .83$), which indicated reliability in our metric [Cronbach 1951]. Table 3 details the raw preferences indicated by participants.

A one-way repeated measures ANOVA was conducted for each scenario with the comparison (PedVR / Decoupled, PedVR+G / Decoupled, PedVR+G / PedVR) as the within-subjects variable. IBM SPSS Statstics was used for analysis. For the Antipodal Circle, the repeated measures ANOVA with a Greenhouse-Geisser correction indicated a statistically significant difference between the mean preference of each comparison F(1.449, 27.528) = 8.488, $p = .003$. Post hoc tests using a Bonferroni correction showed a significant difference between the PedVR / Decoupled comparison and the PedVR+G / PedVR comparison ($p = .006$). For the Bidirectional Flow scenario, the repeated measures ANOVA with Greenhouse-Geisser correction indicated a statistically significant difference between the mean preference of each comparison F(1.728, 32.826) = 4.216, $p = .003$. Post hoc tests using a Bonferroni correction showed a significant difference between the PedVR / Decoupled comparison and the PedVR+G / PedVR comparison ($p = .003$).

Gaze Behavior: Additional analysis using occurrence of preference values suggests that the presence of gaze behaviors has a significant impact on the level of preference of a participant. In the Antipodal Circle scenario, the participants preferred PedVR (response < 4) to Decoupled in 41.3% of all responses, with 8.8% indicating strong preference (response = 1). 31.9% of participants indicated no difference. As gaze behaviors were introduced, participants preferred PedVR+G to Decoupled in 56.2% of responses (36% improvement), with 35.6% indicating a strong preference (400% improvement). Only 10% of responses indicated no difference. These results suggest that gaze behaviors provide a substantial improvement to the sense of presence with the virtual agents in the virtual environment.

In the bidirectional scenario, similar trends were observed. Participants preferred PedVR in 42.5% of responses, 12.5% being strong preference, and indicated no difference in 25.6%. With gaze be-

Figure 5: ***User Evaluation.*** *(A) A user wearing the DK-2. (B) The user was asked to move to the red sphere in each scenario. (C) Antipodal circle scenario (D) Bidirectional flow scenario. For both scenarios, the user was presented with three trials, one for each method.*

haviors, participants preferred PedVR+G in 56.9% of responses (33.9% improvement), 30.0% being strong preference (240% improvement), and only 12.5% of participants indicated no difference. This again suggests that the presence of gaze behaviors has a substantial impact on participants' sense of social presence.

Comparing PedVR and PedVR+G, participants preferred PedVR+G in both scenarios. In the Antipodal Circle, 71.3% of responses favored PedVR+G with 51.2% indicating strong preference. In the Bidirectional Flow scenario, 68.2% of responses favored PedVR+G with 36.3% indicating strong preference. Fig 6 illustrates the response distribution for the Antipodal Circle scenario for the three comparisons.

A final observation during experimentation was the presence of vocal utterances during the task in the PedVR+G condition. Several participants apologized to virtual agents upon collision or greeted the virtual agents when the gaze behaviors engaged. The experimenters did not observe this phenomenon in the other conditions. Although anecdotal, these occurrences reflect the observation that non-verbal behaviors such as gaze and gesture have an impact on the perception of social awareness and presence of virtual agents.

7 Conclusion, Limitations & Future Work

We have presented a novel interactive approach, PedVR, for high dimensional trajectory computation by coupling 2D navigation with full body motion synthesis and combine with gaze computation. Our approach provides the user the ability to interact with the virtual crowd in immersive virtual environments. The virtual agents compute smooth, collision-free trajectories to avoid the user as well as other virtual agents. In addition to collision avoidance, the virtual agents are capable of exhibiting gaze and gestural behaviors to increase the believability of the virtual experience. The results of a within-subjects user study demonstrate a significant preference for our approach, PedVR, compared to existing decoupled crowd simulation algorithms. Furthermore, our results indicate a 4-fold increase in preference for our method with the introduction of gaze.

Our approach is a first step towards immersive crowd simulations and has some limitations. First, the agents are restricted in their ability to generate appropriate gestural responses and communications. Second, the user is limited to using a keyboard and mouse to move in the virtual environment. Previous studies show that real walking increases a subject's sense of presence [Slater et al. 1995], but this requires a larger physical space with accurate tracking. Third, our user evaluation is based on subjective questionnaires and does not take into account physiological responses or breaks-in-presence as a metric for measuring presence.

There are many avenues of future work. Besides overcoming the limitations, we would like to incorporate full body tracking and develop appropriate gestural recognition and response mechanisms to allow for a more behaviorally rich human-like interaction. We would also like to conduct a more expansive user evaluation to study the effectiveness of our approach and use it for different applications. In addition, recent work has explored the use of elliptical 2D agents as opposed to disc agents [Best et al. 2016]. Elliptical agents can more readily engage in shoulder turning and respond more appropriately to personal space considerations. We will investigate the use of such elliptical agents in future experimentation.

Acknowledgements

This research is supported in part by ARO grant W911NF-16-1-0085 and a grant from Boeing.

References

Adams, R. B., and Kleck, R. E. 2003. Perceived gaze direction and the processing of facial displays of emotion. *Psychological Science 14*, 6, 644–647.

Bailenson, J. N., Beall, A. C., Loomis, J., Blascovich, J., and Turk, M. 2005. Transformed social interaction, augmented gaze, and social influence in immersive virtual environments. *Human communication research 31*, 4, 511–537.

Best, A., Narang, S., and Manocha, D. 2016. Real-time reciprocal collision avoidance with elliptical agents. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 298–305.

Bönsch, A., Weyers, B., Wendt, J., Freitag, S., and Kuhlen, T. W. 2016. Collision avoidance in the presence of a virtual agent in small-scale virtual environments. In *IEEE Symposium on 3D User Interfaces*, 145–148.

Bruderlin, A., and Calvert, T. 1993. Interactive animation of personalized human locomotion. In *the Graphics Interface*, 17–23.

Cassell, J. 2001. Embodied conversational agents: representation and intelligence in user interfaces. *AI magazine 22*, 4, 67.

Cirio, G., Olivier, A., Marchal, M., and Pettré, J. 2013. Kinematic evaluation of virtual walking trajectories. *Visualization and Computer Graphics, IEEE Transactions on 19*, 4, 671–680.

Cronbach, L. J. 1951. Coefficient alpha and the internal structure of tests. *psychometrika 16*, 3, 297–334.

Feng, A. W., Xu, Y., and Shapiro, A. 2012. An example-based motion synthesis technique for locomotion and object manipulation. In *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, ACM, 95–102.

Gallup, A. C., Chong, A., Kacelnik, A., Krebs, J. R., and Couzin, I. D. 2014. The influence of emotional facial expressions on gaze-following in grouped and solitary pedestrians. *Scientific reports 4*.

Scenario	Left	Right	1	2	3	4	5	6	7	mean	SD
Antipodal Circle	PedVR+G	Decoupled	57	28	5	16	13	17	24	3.29	± 1.66
	PedVR	Decoupled	14	20	32	51	18	19	6	3.75	± 0.98
	PedVR+G	PedVR	82	18	14	14	6	17	9	2.57	± 1.22
Bidirectional	PedVR+G	Decoupled	48	28	15	20	11	17	21	3.33	± 1.59
	PedVR	Decoupled	20	15	33	41	21	26	4	3.76	± 1.16
	PedVR+G	PedVR	58	34	17	16	18	12	5	2.73	± 1.19

Table 3: ***User Study Responses.*** *This table shows the response frequencies for each comparison and each scenario for the Social Presence questionnaire. For each question, participants rated their preference from 1 to 7 between the left and right methods being evaluated where 1 indicates the left method is "much better" and 7 indicates the right method is "much better". This table also shows the mean social presence score for each comparison for each scenario. In each case, PedVR was preferred to the decoupled method, and the introduction of gaze behaviors increased the proportion of responses that strongly favored our method.*

(A) PedVR vs Decoupled *(B) PedVR+G vs Decoupled* *(C) PedVR+G vs PedVR*

Figure 6: ***Comparison of Preference in the Antipodal Circle scenario.*** (A) PedVR is preferred to Decoupled with 31.9% of participants reporting no difference. *(B) PedVR+G is preferred to Decoupled with 35.6% of participants indicating strong preference.* (C) PedVR+G is preferred to PedVR with 51.2% of participants indicating strong preference.

GARAU, M., SLATER, M., PERTAUB, D.-P., AND RAZZAQUE, S. 2005. The responses of people to virtual humans in an immersive virtual environment. *Presence: Teleoperators and Virtual Environments 14*, 1, 104–116.

HELBING, D., FARKAS, I., AND VICSEK, T. 2000. Simulating dynamical features of escape panic. *Nature 407*, 487–490.

JAIN, S., YE, Y., AND LIU, C. K. 2009. Optimization-based interactive motion synthesis. *ACM Trans. Graph. 28*, 1 (Feb.), 10:1–10:12.

KAPADIA, M., MARSHAK, N., SHOULSON, A., AND BADLER, N. I. 2014. ADAPT: The agent development and prototyping testbed. *IEEE Transactions on Visualization and Computer Graphics 20*, 7 (July), 1035–1047.

KARAMOUZAS, I., SKINNER, B., AND GUY, S. J. 2014. Universal power law governing pedestrian interactions. *Physical review letters 113*, 23, 238701.

KENNEDY, R. S., LANE, N. E., BERBAUM, K. S., AND LILIENTHAL, M. G. 1993. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology 3*, 3, 203–220.

KIEFER, A. W., RHEA, C. K., AND WARREN, W. H. 2013. Vr-based assessment and rehabilitation of functional mobility. In *Human Walking in Virtual Environments*. Springer, 333–350.

KIM, S., BERA, A., BEST, A., CHABRA, R., AND MANOCHA, D. 2016. Interactive and adaptive data-driven crowd simulation. *Proc. of IEEE VR.*

KOVAR, L., GLEICHER, M., AND PIGHIN, F. 2002. Motion graphs. In *ACM transactions on graphics (TOG)*, vol. 21, ACM, 473–482.

LLOBERA, J., SPANLANG, B., RUFFINI, G., AND SLATER, M. 2010. Proxemics with multiple dynamic characters in an immersive virtual environment. *ACM Transactions on Applied Perception (TAP) 8*, 1, 3.

NARANG, S., RANDHAVANE, T., BEST, A., SHAPIRO, A., AND MANOCHA, D., 2016. Fbcrowd: Interactive multi-agent simulation with coupled collision avoidance and human motion synthesis. Available at http://gamma.cs.unc.edu/pedvr.

PARK, C., BEST, A., NARANG, S., AND MANOCHA, D. 2015. Simulating high-dof human-like agents using hierarchical feedback planner. In *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology*, ACM, 153–162.

PELECHANO, N., STOCKER, C., ALLBECK, J., AND BADLER, N. 2008. Being a part of the crowd: towards validating vr crowds using presence. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems*, 136–142.

PERTAUB, D.-P., SLATER, M., AND BARKER, C. 2002. An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence: Teleoperators and virtual environments 11*, 1, 68–78.

PETERS, C., PELACHAUD, C., BEVACQUA, E., MANCINI, M., AND POGGI, I. 2005. A model of attention and interest using gaze behavior. In *International Workshop on Intelligent Virtual Agents*, Springer, 229–240.

REN, Z., GAI, W., ZHONG, F., PETTRÉ, J., AND PENG, Q. 2013. Inserting virtual pedestrians into pedestrian groups video with behavior consistency. *The Visual Computer 29*, 9, 927–936.

RIVALCOBA, J., GYVES, O., RUDOMIN, I., AND PELECHANO, N. 2014. Coupling pedestrians with a simulated virtual crowd. In *International Conference on Computer Graphics and Applications (GRAPP)*.

ROMANO, D. M., AND BRNA, P. 2001. Presence and reflection in training: Support for learning to improve quality decision-making skills under time limitations. *CyberPsychology & Behavior 4*, 2, 265–277.

ROTHBAUM, B. O., HODGES, L. F., READY, D., AND ALARCON, R. D. 2001. Virtual reality exposure therapy for vietnam veterans with posttraumatic stress disorder. *The Journal of clinical psychiatry 62*, 8, 1–478.

SCHADSCHNEIDER, A. 2002. Cellular automaton approach to pedestrian dynamics - theory. *Pedestrian and Evacuation Dynamics*, 75–86.

SHAPIRO, A. 2011. Building a character animation system. In *Motion in Games*, Springer Berlin / Heidelberg, J. Allbeck and P. Faloutsos, Eds., vol. 7060 of *Lecture Notes in Computer Science*, 98–109.

SLATER, M., AND STEED, A. 2000. A virtual presence counter. *Presence 9*, 5, 413–434.

SLATER, M., USOH, M., AND STEED, A. 1995. Taking steps: the influence of a walking technique on presence in virtual reality. *ACM Transactions on Computer-Human Interaction (TOCHI) 2*, 3, 201–219.

SLATER, M., GUGER, C., EDLINGER, G., LEEB, R., PFURTSCHELLER, G., ANTLEY, A., GARAU, M., BROGNI, A., AND FRIEDMAN, D. 2006. Analysis of physiological responses to a social situation in an immersive virtual environment. *Presence: Teleoperators and Virtual Environments 15*, 5, 553–569.

SNOOK, G. 2000. Simplified 3D movement and pathfinding using navigation meshes. In *Game Programming Gems*. Charles River, Hingham, Mass., ch. 3, 288–304.

THIEBAUX, M., LANCE, B., AND MARSELLA, S. C. 2009. Real-Time Expressive Gaze Animation for Virtual Humans. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*.

TREUILLE, A., COOPER, S., AND POPOVIĆ, Z. 2006. Continuum crowds. In *Proc. of ACM SIGGRAPH*, 1160–1168.

ULICNY, B., AND THALMANN, D. 2001. *Crowd simulation for interactive virtual environments and VR training systems*. Springer.

VAN DEN BERG, J., PATIL, S., SEWALL, J., MANOCHA, D., AND LIN, M. 2008. Interactive navigation of multiple agents in crowded environments. In *Proc. Symp. Interact. 3D Graph. Game.*, 139–147.

VAN DEN BERG, J., GUY, S. J., LIN, M., AND MANOCHA, D. 2011. Reciprocal n-body collision avoidance. In *Inter. Symp. on Robotics Research*, 3–19.

VON DER PÜTTEN, A. M., KRÄMER, N. C., AND GRATCH, J. 2009. 1 who´ s there? can a virtual agent really elicit social presence?

WELBERGEN, V. H., BASTEN, V. B., EGGES, A., RUTTKAY, Z., AND OVERMARS, M. 2010. Real time character animation: A trade-off between naturalness and control. *Computer Graphics Forum 29*, 8.