F2FCrowds: Planning Agent Movements to Enable
Face-to-Face Interactions

Tanmay Randhavane, Aniket Bera, Dinesh Manocha

UNC Chapel Hill

E-mail: tanmay@cs.unc.edu, ab@cs.unc.edu, dm@cs.unc.edu

Abstract. We present an approach for multi-agent navigation that facilitates face-to-face interaction in virtual crowds. We describe a model of approach behavior for virtual agents that includes a novel interaction velocity prediction (IVP) algorithm. This algorithm is combined with human body motion synthesis constraints and facial actions to improve the behavioral realism of virtual agents. We combine these techniques with full-body crowd simulation and evaluate their benefits by conducting a user study using immersive hardware. Our results indicate that such techniques enabling face-to-face interactions can improve the sense of presence felt by the user. The virtual agents using these algorithms also appear more responsive and are able to elicit more reaction from the users.
Figure 1: **F2FCrowds**: Our algorithm enables the real user (in blue, wearing an HMD) to have face-to-face interactions with virtual agents. The virtual agents are responsive and exhibit head movements, gazing and gesture behaviors.

### 1. Introduction

In many applications, it is important to simulate the behavior of virtual humans and crowds. It is well known that adding virtual agents or avatars into simulated worlds can improve the sense of immersion (Llobera, Spanlang, Ruffini, & Slater, 2010; Nuria Pelechano, Stocker, Allbeck, & Badler, 2008; Slater et al., 2006). The use of virtual characters and associated environments is widely adopted in training and rehabilitation environments (Ulicny & Thalmann, 2001). Other applications include treatment of crowd phobias and social anxiety using VR therapy (Pertaub, Slater, & Barker, 2002), architectural flow analysis and evacuation planning (Cassol, Oliveira, Musse, & Badler, 2016; Haworth et al., 2016), etc.

There is considerable work on evaluating the sense of presence and immersion in VR based on the behaviors, interactions, and movements of virtual agents. Many researchers have concluded that the social presence of virtual agents depends on the realism of their behavior (Blascovich et al., 2002) and the nature of their interactions (Guadagno, Blascovich,
Bailenson, & McCall, 2007; Kyriakou, Pan, & Chrysanthou, 2015). Recent advances in artificial intelligence (including natural language processing and computer vision, along with development of embodied conversational agents) are helping to generate realistic interaction scenarios. Other work includes the development of techniques to simulate gazing, collision-avoidance movements, head turning, facial expressions, and other gestures (Grillon & Thalmann, 2009; Nummenmaa, Hyöna, & Hietanen, 2009).

One of the main social interactions is face-to-face (F2F) interaction, which is typically carried out without the use of any mediating technology (Crowley & Mitchell, 1994). This broad area has been studied in social sciences for more than a century. There is a recent interest in integrating virtual reality technology into social media, and F2F communication is an important component of such a system. Different sensory organs play an important role in these interactions, which may include eye contact or two agents facing or talking to each other in close proximity. As a result, there are many challenges in terms of developing such interaction capabilities between virtual agents.

**Main Results:** We address the problem of computing the movements or trajectories to enable F2F interactions between a real user and a virtual agent who is part of a virtual crowd. This includes automatically computing collision-free trajectories that enable such agents to come close to each other for F2F communications. Satake et al. (2009) developed a model of approach behavior for robots having F2F interactions with people who are walking. Their model was based on the idea that human interactions can be classified based on social and public distance (Hall, 1966). Motivated by these ideas, we develop a model of approach behavior for virtual agents that models their movement for F2F communication. We present a novel navigation algorithm, *Interaction Velocity Prediction (IVP)*, which predicts whether the avatar of a real user is trying to approach a virtual agent for F2F interaction. IVP is combined with 2D multi-agent simulation to compute collision-free trajectories. In order
to generate plausible full-body simulations, we also integrate the velocity computation with human motion synthesis to generate upper body movements such as gazing and nodding. Overall, our approach (F2FCrowds) can generate smooth and natural-looking trajectories for each agent. We use our algorithms to simulate the movement of tens of virtual agents in complex indoor and outdoor environments at interactive rates.

In order to evaluate the benefits of our algorithms, we performed a user evaluation in an immersive environment where a real user interacted with the virtual agents. In particular, we compared the following three approaches for simulating virtual agents and crowds:

- **PedVR**: A crowd simulation algorithm that uses coupled 2D navigation and full-body motion synthesis (Narang, Best, Randhavane, Shapiro, & Manocha, 2016).

- **F2FCrowds without Upper Body Behaviors**: The 2D multi-agent simulation in PedVR is combined with a model of approach behavior to compute agent movements.

- **F2FCrowds with Upper Body Behaviors**: In addition to the model of approach behavior, we also simulate gazing and head movements for F2F interactions.

We have conducted a within-subject user study with 15 subjects and evaluated the results in two scenarios using an Oculus Rift HMD. Our studies are designed based on the prior work of Garau, Slater, Pertaub, and Razzaque (2005), Narang, Best, et al. (2016), Nuria Pelechano et al. (2008), which evaluated the level of presence based on the behavior and interactions with the virtual agents in a crowd. We observe a statistically significant preference for our new algorithms in F2FCrowds when compared to PedVR. Our algorithm increased the sense of presence felt by the users. When using our algorithms, the virtual agents appeared more responsive and were able to elicit more reaction from the users. Our results for the sense of presence question show that 40% of the participants preferred F2FCrowds (without upper body behaviors) over PedVR, whereas only 3.33% participants preferred PedVR and the rest remained neutral. Participants felt virtual agents were more aware for F2FCrowds (without
upper body behaviors) in 60% of the responses, whereas only 3.33% felt that way for PedVR. Our methods (without upper body behaviors) elicited more reaction from the user in 53.33% cases, whereas PedVR elicited more reaction in 10% of the cases. The addition of upper body behaviors to F2FCrowds also showed a significant improvement in the performance.

The rest of the paper is organized as follows. We briefly present prior work on crowd simulation and interactions in Section 2. In Section 3, we provide an overview of our algorithm. We describe the model of approach behavior for virtual agents and the novel velocity computation algorithms in Section 4. In Section 5, we provide the implementation details and highlight our algorithm’s performance on different benchmarks. We describe the details of our user evaluation in Section 6.

2. Related Work

In this section, we give an overview of prior work on face-to-face interactions, crowd simulation for VR and interaction with virtual agents in a virtual environment.

2.1. Face-to-Face Interactions

Face-to-face interactions have been studied in psychology, sociology, and robotics. Satake et al. (2009) presented an algorithm to enable a robot to have F2F interactions with people who are walking. Gonçalves and Perra (2015) studied empirical characteristics of face-to-face interaction patterns and novel techniques to discover mesoscopic structures in these patterns. There is work on investigating F2F interactions in terms of capabilities to understand and generate natural language in combination with non-verbal signals and social management (Bonaiuto & Thórisson, 2008; Cassell, Vilhjálmsson, & Bickmore, 2001; Heylen et al., 2011; Jonsdottir, Thorisson, & Nivel, 2008; Kopp, Stocksmeier, & Gibbon, 2007). In the new field of Social Signal Processing, studies are being performed to understand and model social
interactions and to provide computers with these capabilities while interacting with humans (Pantic et al., 2011; Vinciarelli, Pantic, & Bourlard, 2009). In this paper, we attempt to provide the users of virtual reality the ability to have F2F interactions with virtual agents in virtual crowds. Our approach provides a platform to implement the aforementioned models in the context of F2F interactions in virtual crowds.

2.2. Crowd and Multi-Agent Simulation

A significant amount of research has been done in multi-agent and crowd simulation. In this paper, we mainly limit ourselves to a class of algorithms that decomposes the trajectory or behavior computation for each agent into two parts: global planning and local navigation (Helbing & Molnar, 1995; Kapadia & Badler, 2013; Ondřej, Pettré, Olivier, & Donikian, 2010; Reynolds, 1999; Van Den Berg, Guy, Lin, & Manocha, 2011). The global planner computes a path for each agent in the environment towards its intermediate goal position. The local navigation algorithms modify these paths so that the agents can avoid collisions with dynamic obstacles or other pedestrians in the environment. Some of these methods also account for a pedestrian’s personality (Guy, Kim, Lin, & Manocha, 2011; N. Pelechano, Allbeck, & Badler, 2007) or use cognitive techniques (Funge, Tu, & Terzopoulos, 1999).

2.3. Interaction With Virtual Agents

There is extensive literature on simulating realistic behaviors, movements, and interactions with virtual agents in VR (Magnenat-Thalmann & Thalmann, 2005). In this paper, we restrict ourselves to modeling some of the interactions between real and virtual agents when they are in close proximity. Kyriakou et al. (2015) showed that basic interaction increases the sense of presence, though they did not explicitly model users’ intent to participate. Bailenson, Blascovich, Beall, and Loomis (2001) concluded that there is an inverse relationship between gazing and personal space. Nuria Pelechano et al. (2008) showed that pushing-based
interaction increases the sense of presence in a virtual environment. Bonsch, Weyers, Wendt, Freitag, and Kuhlen (2016) described a gaze-based collision avoidance system and interactions for small-scale virtual environments. Hu, Adeagbo, Interrante, and Guy (2016) presented a system where virtual agents exhibit head turning behavior but do not explicitly model face-to-face interaction. There is also considerable work on individualized avatar-based interactions (Nagendran, Pillat, Kavanaugh, Welch, & Hughes, 2014). Our approach to enabling F2F interactions between real and virtual agents is complimentary to most of these methods and can be easily combined with them.

3. Overview

In this section, we introduce our notation and give an overview of our approach for crowd simulation to enable F2F interactions.

3.1. Notation

Our approach uses a multi-agent simulation algorithm that computes the trajectory of each agent using a combination of global planning and local navigation. We assume that the
environment consists of a real user, represented by its avatar, and virtual users or agents. The real user is walking or navigating in an immersive setting in an environment with many virtual agents, avoiding collisions and attempting to interact with the virtual agents. One of our goals is to compute collision-free and plausible trajectories for each virtual agent to enable F2F interactions with the user. We model this using a novel model of approach behavior based on Hall’s idea of social and public distance (Hall, 1966). This model makes use of the novel IVP algorithm to predict whether the avatar of the real user is trying to approach a virtual agent to perform F2F interactions.

We represent each agent using a high-DOF articulated model and compute upper and lower body motions. The state of an agent $i$ is represented by $q_i$ and is the union of the position of the root joint and the states of all the joints of the high-DOF character. In terms of 2D multi-agent navigation, we represent an agent $i$ as a circle of radius $r_i$ at a position $\vec{p}_i$, which is the 2D position of the root joint of the agent. At any time, a virtual agent $i$’s current velocity is represented by $\vec{v}_i$ and its preferred velocity and preferred orientation are represented by $\vec{v}_i^p$ and $o_i^p$, respectively. The preferred velocity and orientation are based on the intent of the virtual agent. We use $M$ to denote the set of available upper body behaviors and $m_i$ to denote the current upper body behavior of agent $i$. The 3D point on the face of the user avatar at which the virtual agent $i$ is currently gazing is denoted by $\vec{g}_i$ (i.e. the gazing point).

We represent the user’s avatar with the subscript $u$. Let $S$ be the simulator state, which is the union of the states of all the entities in the scene, including obstacles and agents.

Figure 2 provides an outline of our interactive crowd simulation pipeline. We use a game engine to gather the user’s input, which is then used by our multi-agent simulation system. The 2D multi-agent system uses a layered 2D navigation algorithm. The first layer corresponds to the model of approach behavior and global planning, which computes the preferred velocity $\vec{v}_i^p$ and preferred orientation $o_i^p$ of each virtual agent $i$. The second layer, local navigation,
computes the collision-free velocities \( \vec{v}_c^i \) for each virtual agent \( i \). The computed velocity \( \vec{v}_c^i \) and upper body behavior motions are passed to the motion synthesis module, which computes the state \( q_i \) for each virtual agent \( i \).

4. Model of Approach Behavior

According to Hall (1966), the term ”Public Distance” refers to the distance at which people can give a speech and ”Social Distance” characterizes the distance at which people can talk to each other. Hall (1966) classified human interactions based on the idea of these public and social distances. Based on these ideas, Satake et al. (2009) proposed a model of approach behavior with which a robot can initiate conversation with people who are walking. In order to have a F2F interaction, a robot should find a person with whom to talk, start approaching that person at a public distance, and initiate the conversation at a social distance. Therefore, Satake et al. (2009) defined ”approach behavior” as a sequence of the following activities: (1) selecting a target, (2) approaching the target at public distance, and (3) initiating conversation at social distance. We use similar ideas and propose a model of approach behavior for virtual agents (Figure 3) that models how the virtual agent should approach the user. The model is a sequence of the following activities: (1) identifying the intent of interaction of the user, (2) approaching at public distance \( (d_p) \), and (3) initiating communication at social distance \( (d_s) \).

4.1. Identifying the Intent of Interaction of the User

A user may wish to interact with a virtual agent that is currently farther than public distance \( (d_p) \). In order to have an interaction, the user will attempt to be within the social distance \( (d_s) \) of the virtual agent by moving towards the virtual agent. The virtual agent then must be able to identify this intent of interaction of the user in order to have a F2F interaction. To achieve this, we use a novel algorithm, **Interaction Velocity Prediction (IVP)**, that predicts whether
Figure 3: **Model of Approach Behavior**: We define the model of approach behavior for virtual agents as a sequence of three activities based on the distance between the user and the virtual agent ($d$): (1) identifying the intent of interaction of the user, (2) approaching at public distance ($d_p$), and (3) initiating communication at social distance ($d_s$).

Figure 4: **Interaction Velocity Prediction** Given the current position ($\mathbf{p}_i^c$) and preferred velocity of the virtual agent ($\mathbf{v}_i^o$) and the current position of the user agent ($\mathbf{p}_u^c$), our IVP algorithm predicts the velocity ($\mathbf{v}_{u}^{\text{ivp}}$) of the user agent to intercept the virtual agent in time $t_{\text{min}}$. If the user’s predicted velocity $\mathbf{v}_{u}^{\text{pred}}$ satisfies the constraint $\mathbf{v}_{u}^{\text{ivp}} \cdot \mathbf{v}_{u}^{\text{pred}} \geq \theta_v$, it will result in F2F communication.

or not a user is trying to interact with a virtual agent.

Given the current position $\mathbf{p}_i^c$ and preferred velocity $\mathbf{v}_i^o$ of a virtual agent $i$ and the current
position of the user agent $\vec{p}_u$, our IVP algorithm $\text{IVP}_i : \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}^2 \times \mathbb{R}$ determines the velocity ($\vec{v}_{ivp}^u$) that the real user should follow to intercept the virtual agent in time $t_{\text{min}}$.

If the public distance is given by $d_p$, then the time of interception $t$ can be given as:

$$\left\| \vec{p}_u - \vec{p}_i \right\| \leq d_p. \quad (1)$$

Assuming that the user agent has the velocity $\vec{v}_{ivp}^u$ and the virtual agent has the average velocity $\vec{v}_i$,

$$\left\| (\vec{p}_u + \vec{v}_{ivp}^u t) - (\vec{p}_i + \vec{v}_i t) \right\| \leq d_p, \quad (2)$$

$$\left\| \vec{v}_{ivp}^u - (\vec{v}_i - \frac{(\vec{p}_u - \vec{p}_i)}{t}) \right\| \leq \frac{d_p}{t}. \quad (3)$$

We solve the above equation for interaction velocity $\vec{v}_{ivp}^u$, i.e. when $t$ is minimized. We also take into account motion and dynamic constraints of the agent and put a limit on the maximum speed:

$$\left\| \vec{v}_{ivp}^u \right\| \leq v_{\text{max}}, \quad (4)$$

where $v_{\text{max}}$ is the maximum speed of the user agent. Simplifying these two equations leads to a $4^{\text{th}}$ order polynomial. Therefore, we calculate $\vec{v}_{ivp}^u$ such that the center of the circular virtual agent and circular user agent coincide, i.e. :

$$\vec{v}_{ivp}^u = \vec{v}_i - \frac{(\vec{p}_u - \vec{p}_i)}{t}. \quad (5)$$

Substituting this expression in Equation (4) results in

$$\left\| \vec{v}_i - \frac{(\vec{p}_u - \vec{p}_i)}{t} \right\| \leq v_{\text{max}} \quad (6)$$

$$(\vec{v}_{ix}^u t - (\vec{p}_{ux} - \vec{p}_{ix}))^2 + (\vec{v}_{iy}^u t - (\vec{p}_{uy} - \vec{p}_{iy}))^2 \leq (v_{\text{max}} t)^2. \quad (7)$$
We simplify Equation (7) as \( at^2 + bt + c \leq 0 \), where

\[
a = (\vec{v}_{ix}^o)^2 + (\vec{v}_{iy}^o)^2 - v_{max}^2, \tag{8}
\]

\[
b = -2((\vec{p}_{ux}^c - \vec{p}_{ix}^c)\vec{v}_{ix}^o + (\vec{p}_{uy}^c - \vec{p}_{iy}^c)\vec{v}_{iy}^o), \tag{9}
\]

\[
c = (\vec{p}_{ux}^c - \vec{p}_{ix}^c)^2 + (\vec{p}_{uy}^c - \vec{p}_{iy}^c)^2. \tag{10}
\]

We assume that the agent’s speed will be lower than the user’s speed (otherwise the user agent will never be able to intercept the virtual agent), \( a \leq 0 \). Since \( c > 0 \), \( t_{min} \) is the larger root of the equation \( at^2 + bt + c = 0 \) and the interaction velocity is:

\[
\vec{v}_{ivp}^u = \vec{v}_i^o - \frac{(\vec{p}_u^c - \vec{p}_i^c)}{t_{min}} \tag{11}
\]

### 4.1.1. Computation of Preferred Velocity

We use IVP (Figure 3) to compute the interaction velocity \( \vec{v}_{ivp}^u \) that the user will have in order to have F2F interactions with a virtual agent \( i \) at time \( t_{min} \). Based on the user’s position from the past few frames, we can predict the velocity of the user based on some motion model and denote it as \( \vec{v}_{pred}^u \). In this case, the virtual agent \( i \) will have F2F interaction with the user if:

\[
\vec{v}_{ivp}^u \cdot \vec{v}_{pred}^u \geq \theta_v, \tag{12}
\]

where \( \theta_v \) is a small threshold.

The preferred velocity \( \vec{v}_i^o \) for a virtual agent \( i \) is then computed as follows:

\[
\vec{v}_i^o = v_{pref} \ast \frac{(\vec{p}_u^c - \vec{p}_i^c)}{\| (\vec{p}_u^c - \vec{p}_i^c) \|}, \tag{13}
\]

where \( \vec{p}_i^c \) and \( \vec{p}_u^c \) are the current positions of the virtual agent \( i \) and user agent \( u \), respectively, and \( v_{pref} \) is the preferred natural speed of the virtual agent.
4.2. Approaching at Public Distance

At public distance, the virtual agent and the user can acknowledge each other. The virtual agent achieves this by slowing down and gazing at the user. We define a boolean function $\text{approach}_p()$ to denote the conditions when a virtual agent $i$ decides to approach the real user at public distance:

$$\text{approach}_p() = (d_s < \|\vec{p}_c^u - \vec{p}_c^i\| < d_p) \land (\vec{o}_c^u \cdot \|\vec{p}_c^u - \vec{p}_c^i\| > o_{\text{thresh}}),$$

where $\vec{o}_c^u$ is the 2D orientation vector of the user, $o_{\text{thresh}}$ is a pre-determined threshold, and $d_s$ and $d_p$ are the social and public distances, respectively.

When $\text{approach}_p()$ evaluates to true, the virtual agent slows down to allow a friendly approach and its preferred velocity is given by:

$$\vec{v}_i^v = k \cdot v_{\text{pref}} \cdot (\vec{p}_c^u - \vec{p}_c^i),$$

Here, $0 < k < 1$ is a pre-determined constant. Notice that the virtual agent’s speed is directly proportional to the distance between the agent and the user to slow down the virtual agent as it approaches the user.

4.2.1. Gazing  In addition to computing the appropriate velocities, it is also important to exhibit appropriate upper body movements and behaviors for F2F communications. Gazing plays an important role in conveying the intent of interaction and it is important for virtual agents to maintain eye contact with the user while approaching. Therefore, the virtual agents gaze at the eyes of the user’s 3D avatar (Figure 6) whenever $\text{approach}_p()$ evaluates to true. We do this by setting the gazing point $g_i$ of the virtual agent $i$ to the position of the eye of the user’s 3D avatar.
4.3. Initiating Communication at Social Distance

Social distance is the distance at which humans typically have face-to-face interactions in social scenarios (Hall, 1966). Therefore, when the distance between the real user and the virtual agent is less than social distance, the virtual agent stops and attempts to have a communication with the user as denoted by the boolean function $approach_s()$:

$$approach_s() = (\|\vec{p}_u^c - \vec{p}_i^c\| < d_s) \land \left(\frac{\vec{o}_u \cdot (\vec{p}_u^c - \vec{p}_i^c)}{\|\vec{p}_u^c - \vec{p}_i^c\|} > o_{thresh}\right),$$

where $\vec{o}_u$ is the 2D orientation vector of the user, $o_{thresh}$ is a pre-determined threshold, and $d_s$ is the social distance.

4.3.1. Head Movements Head movements play an important part in F2F interactions (McClave, 2000). Therefore, our virtual agents exhibit head movements like nodding, shaking, and tossing their heads to communicate with the user (Figure 6). During the communication, the virtual agent performs head movements at randomized time intervals ranging from $6 - 10$ seconds (based on Hadar, Steiner, and Clifford Rose (1985)). The head movement is chosen at random from the set of motions $M = \{nod, toss, shake\}$. Since nod implies a positive intent of interaction, the first head movement is always chosen to be a nod. The virtual agent concludes that the communication is over when $approach_s()$ evaluates to false.

4.4. Navigation

The model of approach behavior discussed so far determines whether or not a virtual agent is a part of an interaction and then calculates its preferred velocity. All the other agents that are not part of any interaction follow goal-directed behavior. We use algorithms described in Narang, Randhavane, Best, Shapiro, and Manocha (2016) to plan the movement of these agents and
Figure 5: **F2F Communications** Our approach enables F2F communications between the avatar of the real user and the virtual agent.

Figure 6: **Gestures**: Our virtual agents exhibit head movements and gazing. Appropriate head movements are chosen from the set of movements including (a) nod (vertical head movement), (b) shake (horizontal head movement), and (c) toss (sideways head movement). (d) Virtual agents also gaze at the user agent to establish eye contact.

determine their preferred velocity. We use constraint modeling from Narang, Randhavane, et al. (2016) to model the collision avoidance constraints and modify the preferred velocity of each agent \( i \) to get the current velocity \( \vec{v}_i \).

5. Implementation and Results

We have implemented our system on a Windows 10 desktop PC with Intel Xeon E5-1620 v3 in parallel on 4 cores and 16 GB of memory. We use Menge (Curtis, Best, & Manocha, 2016)
Figure 7: **Benchmarks:** We highlight the performance of our algorithm on three benchmarks. (a) A shopping mall scene shows virtual agents walking in a mall. (b) The user agent travels in a crossing scenario with multiple virtual agents who gaze at the user’s avatar. (c) Virtual agents explore a tradeshow scenario and acknowledge the user avatar’s presence with eye contact. We are able to simulate tens of agents at interactive rates and evaluate the benefits of F2F interactions.

as our multi-agent simulation library that computes the 2D trajectory for each agent. We have modified the global planning and local navigation algorithms based on the components described by Narang, Randhavane, et al. (2016). A game engine (Unreal Engine 4) serves as an interface to the user and as a rendering framework. We use Smartbody (Shapiro, 2011) to synthesize motion of virtual agents and to provide the joint angles to simulate the motions corresponding to various gestures or upper body movements.

Table 1 highlights the performance of our system on the following benchmark scenarios (Figure 7):

- **Shibuya Crossing** A busy crossing scenario. We initialize the agents at different positions of the intersection. The goal positions are assigned using a probability distribution. After reaching the goals, each agent waits for a few seconds and then moves towards the next goal.

- **Shopping Mall** Virtual agents walk in a shopping mall. They walk to the shops and exhibit head movements (nod or shake for approval or disapproval, respectively) and gazing behaviors at the shops.

- **Tradeshow** Virtual agents walk up to the booths in a tradeshow and exhibit head movements.
## Benchmarks

<table>
<thead>
<tr>
<th>Benchmark</th>
<th>Agents</th>
<th>Average Frame Update Time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PedVR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F2F</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F2F+G</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mall</td>
<td>24</td>
<td>19 20 24</td>
</tr>
<tr>
<td>Shibuya</td>
<td>32</td>
<td>18 18 20</td>
</tr>
<tr>
<td>Tradeshow</td>
<td>30</td>
<td>21 22 26</td>
</tr>
<tr>
<td>Circle</td>
<td>8</td>
<td>9 9 11</td>
</tr>
<tr>
<td>Bidirectional</td>
<td>8</td>
<td>8 8 10</td>
</tr>
</tbody>
</table>

Table 1: **Average frame update time:** In the absence of upper body movements, F2FCrowds with IVP does not have significant overhead over PedVR. F2FCrowds with gesture can simulate 30+ virtual agents at 40-60 FPS.

Our system can simulate 30+ agents at approximately 40-60 FPS.

## 6. User Evaluation

In this section, we describe our user study, which was conducted to evaluate our new algorithms that enable F2F interactions. We performed a within-users study showing the advantages of our model of approach behavior and upper body motion generation.

### 6.1. Study Goals and Expectations

This study was aimed at measuring the advantage of our model of approach behavior for virtual agents over an interactive system. We compared our algorithms with PedVR (Narang, Best, et al., 2016), which is a coupled crowd simulation algorithm that combines 2D multi-agent navigation and 3D full body synthesis to simulate full-body crowds with plausible movements but does not model any approach behaviors. We expected to find that participants felt it easier to have F2F interactions with our algorithm and that these interactions also benefited from the addition of head movements and gazing behaviors. In particular, we propose the following hypotheses:

- **Hypothesis 1:** Addition of model of approach behavior and upper body motion generation increases the sense of presence felt by the user.
Figure 8: **User Interacting with the Virtual Agent** Participants approached virtual agents and attempted to have a F2F interaction.

- Hypothesis 2: Users do not have to make extra effort to avoid the virtual agents after the addition of model of approach behavior and upper body motion generation.
- Hypothesis 3: Virtual agents appear more aware of the user after the addition of model of approach behavior and upper body motion generation.
- Hypothesis 4: Addition of model of approach behavior and upper body motion generation elicits more response from the users.
- Hypothesis 5: Virtual agents appear more responsive after the addition of model of approach behavior and upper body motion.

6.2. Experimental Design

A within-users study was performed in which the participants were asked to participate in two scenarios using an Oculus Rift head mounted display. Participants were standing up and used a joystick for movement in the virtual world (Figure 8). A training scenario was also presented to familiarize the participants with the movement. The participants performed three trials of each scenario and answered a questionnaire at the end of each scenario. We conducted the study in a laboratory setting.
6.2.1. Evaluated Methods  In the study, participants compared three different interaction enabling algorithms:

- **PedVR**: We used the coupled crowd simulation method PedVR as the baseline. The virtual characters in this system used coupled 2D navigation and full-body motion synthesis (Narang, Best, et al., 2016). Gazing and head movements were not included in this algorithm.

- **F2FCrowds**: Our model of approach behavior was used to enable F2F interactions between the user and virtual agents. Gazing and head movements were not included in this algorithm.

- **F2FCrowdsHead**: In addition to the approach behavior, virtual agents also communicated using gazing and head movements.

6.2.2. Task  Each virtual agent in the scene had a cone-shaped marker over its head. The participants were asked to approach any virtual agent and were informed that when they felt that it was possible to have a F2F interaction with the virtual agent, they should press a button on the joystick. When the button was pressed, the marker on the head of a virtual agent in front of the participant was highlighted for two seconds.

6.2.3. Scenarios  Two scenarios were presented to the participants. The participants performed three trials of each scenario corresponding to each method and answered a questionnaire after each scenario.

- **Circle**: This scenario consisted of 8 virtual agents starting on the perimeter of a circle of radius 16 meters. Their target positions were selected randomly on the perimeter of the circle and the simulation resulted in a high-density area at the center of the circle. Each participant started from a position inside the circle and was asked to approach any virtual
agent in the scene for 45 seconds. Participants performed three trials of this scenario for each of the methods described above.

- **Bidirectional**: 8 virtual agents started from opposite ends of a hallway, with half the agents at either end of the hallway, and traveled between the two ends. The participant (i.e. the real user) started at the middle of the hallway and was asked to approach any virtual agent in the scene for 45 seconds. Participants performed three trials of this scenario for each of the algorithms.

### 6.2.4. Questionnaire

The aim of the user study was to show the benefits of our model of approach behavior for virtual agents and upper body movement. We used a modified version of a well-established questionnaire for social presence (Garau et al., 2005). In particular, we used a subset of the original questions and asked additional questions regarding the participant’s interaction with the virtual agents. The questions were of an Agree/Disagree type and participants noted their preference using a seven-level Likert scale with values labeled "Strongly disagree", "Disagree", "Slightly disagree", "Neutral", "Slightly agree", "Agree", and "Strongly agree". For analysis, we convert the participant responses to a scale of 1 (Strongly Disagree) - 7 (Strongly Agree). We list the questionnaire in Table 2.
Question 1: I had a sense of being in the same space as the characters.
Question 2: I had to make an effort to avoid the characters.
Question 3: The characters seemed to be aware of me.
Question 4: I felt that I should talk/nod/respond to the characters.
Question 5: The characters seemed to respond to my attempts of interaction.
Question 6: The characters seemed to respond even if I did not attempt to interact.

Table 2: Questionnaire. Participants were asked to answer the above questions on a seven-level Agree/Disagree Likert scale.

**Simulator Sickness Index:** We also administered a Simulator Sickness Questionnaire (SSQ) (Kennedy, Lane, Berbaum, & Lilienthal, 1993) before and after the study to ensure that our method did not cause any significant discomfort.

6.3. Participants

The participants in our study comprised of graduate students and staff members and were recruited at a university campus. 15 participants were recruited: 4 females and 11 males, with an average age of 24.4 years with standard deviation of 3.1 years. It was ensured that the participants felt comfortable using an HMD and they were given a high-level overview of the setup before starting the study.

6.4. Procedure

Participants 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. They were given a high-level overview of the setup and informed that they could leave the study at any point in time. The participants performed a training task to get familiarized with movement in the virtual world.

Participants performed the Circle and Bidirectional scenarios in a randomized order that prevented any scenario bias. They performed three trials of each scenario, one trial for each of the target methods, again in a randomized order. After the third trial of each scenario, they
Table 3: **Results of a Friedman Test**: We present the test statistic ($\chi^2$) value and the significance level ($p$) of a Friedman test performed to test for differences between the responses for the three algorithms.

<table>
<thead>
<tr>
<th>Question</th>
<th>Circle $\chi^2$</th>
<th>Circle $p$</th>
<th>Bidirectional $\chi^2$</th>
<th>Bidirectional $p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question 1</td>
<td>10.516</td>
<td>0.005</td>
<td>10.759</td>
<td>0.005</td>
</tr>
<tr>
<td>Question 2</td>
<td>0.167</td>
<td>0.92</td>
<td>0.2</td>
<td>0.905</td>
</tr>
<tr>
<td>Question 3</td>
<td>16.642</td>
<td>0</td>
<td>23.787</td>
<td>0</td>
</tr>
<tr>
<td>Question 4</td>
<td>16.51</td>
<td>0</td>
<td>23.231</td>
<td>0</td>
</tr>
<tr>
<td>Question 5</td>
<td>16</td>
<td>0</td>
<td>22.37</td>
<td>0</td>
</tr>
<tr>
<td>Question 6</td>
<td>13.192</td>
<td>0.001</td>
<td>17.796</td>
<td>0</td>
</tr>
</tbody>
</table>

answered the questionnaire (Table 2).

**6.5. Results and Discussion**

In this section, we present and analyze the participant responses (Figure 9) to the three interaction simulation algorithms described previously. For each scenario, the simulation algorithm is the independent variable and the participant response is the dependent variable. Since our dependent variable is ordinal, we used the **Friedman test** to test for differences between the responses for the three algorithms. We tabulate the test statistic ($\chi^2$) value and the significance level ($p$) in Table 3. Post hoc analysis with **Wilcoxon signed-rank tests** was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < 0.017$. We tabulate the $Z$ statistic and the significance level ($p$) for this test in Table 4 and Table 5. We discuss the results for each question below:

- **Question 1**: Question 1 asked whether participants felt a sense of presence in the virtual environment. For both the scenarios, Friedman test revealed a significant difference in the sense of presence depending on the algorithm used. In the Wilcoxon signed-rank test for the Circle scene, there was no significant difference between the PedVR / F2FCrowds comparison, but significant difference was observed for the F2FCrowds / F2FCrowdsHead and PedVR / F2FCrowdsHead comparisons. For the Bidirectional
PedVR vs F2FCrowds | F2FCrowds vs F2FCrowdsHead | PedVR vs F2FCrowdsHead
---|---|---
**Z** | **p** | **Z** | **p** | **Z** | **p**
Question 1 | -1.807 | 0.071 | -2.456 | 0.014 | -2.395 | 0.017
Question 3 | -2.291 | 0.022 | -2.618 | 0.009 | -2.713 | 0.007
Question 4 | -2.62 | 0.009 | -2.424 | 0.015 | -2.726 | 0.006
Question 5 | -2.625 | 0.009 | -2.647 | 0.008 | -3.097 | 0.002
Question 6 | -2.155 | 0.031 | -2.369 | 0.018 | -2.747 | 0.006

Table 4: **Post hoc test for the Circle scene**: We present the \( Z \) statistic and the significance level (\( p \)) of a post hoc analysis with a Wilcoxon signed-rank test. A Bonferroni correction is applied, resulting in a significance level set at \( p < 0.017 \). Since the results of the Friedman test for Question 2 were not statistically significant, we did not run a post hoc test for this question.

<table>
<thead>
<tr>
<th>PedVR vs F2FCrowds</th>
<th>F2FCrowds vs F2FCrowdsHead</th>
<th>PedVR vs F2FCrowdsHead</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Z</strong></td>
<td><strong>p</strong></td>
<td><strong>Z</strong></td>
</tr>
<tr>
<td>Question 1</td>
<td>-2.46</td>
<td>0.014</td>
</tr>
<tr>
<td>Question 3</td>
<td>-2.549</td>
<td>0.011</td>
</tr>
<tr>
<td>Question 4</td>
<td>-1.652</td>
<td>0.098</td>
</tr>
<tr>
<td>Question 5</td>
<td>-1.839</td>
<td>0.066</td>
</tr>
<tr>
<td>Question 6</td>
<td>-0.855</td>
<td>0.393</td>
</tr>
</tbody>
</table>

Table 5: **Post hoc test for the Bidirectional scene**: We present the \( Z \) statistic and the significance level (\( p \)) of a post hoc analysis with a Wilcoxon signed-rank test. A Bonferroni correction is applied, resulting in a significance level set at \( p < 0.017 \). Since the results of the Friedman test for Question 2 were not statistically significant, we did not run a post hoc test for this question.

In the Circle scene, there was a significant difference between the PedVR / F2FCrowds and PedVR / F2FCrowdsHead comparisons, but significant difference was not observed for the F2FCrowds / F2FCrowdsHead comparison. This proves the **hypothesis 1** that the users feel a sense of presence in the virtual environment after the addition of both the model of approach behavior and upper body motion generation.

- **Question 2**: Question 2 evaluated the effort required to avoid collisions. Participants reported no difference between the three algorithms for both the scenarios as indicated by the Friedman test. This proves the **hypothesis 2** that the users do not have to make extra effort to avoid collisions with the virtual agents after the addition of model of approach
behavior and upper body motion generation. Since the results of the Friedman test were not statistically significant, we did not run a post hoc test for this question.

- **Question 3:** Question 3 evaluated whether the participants felt that the virtual agents were aware of the participant. For both the scenarios, Friedman test revealed a significant difference in the participant responses depending on the algorithm used. For the Circle scene, post hoc tests did not reveal a significant difference between the PedVR / F2FCrowds comparison, but significant difference was observed for the F2FCrowds / F2FCrowdsHead and PedVR / F2FCrowdsHead comparisons. Significant differences were observed for all the three comparisons for the Bidirectional scene proving the hypothesis 3 that virtual agents appear more aware of the user after the addition of model of approach behavior and upper body motion generation.

- **Question 4:** Question 4 evaluated whether the participants felt that they should talk/nod/respond to the characters. Friedman test revealed a significant difference in the participant responses depending on the algorithm used for both the scenarios. For the Bidirectional scene, post hoc tests did not reveal a significant difference between the PedVR / F2FCrowds comparison, but significant difference was observed for the F2FCrowds / F2FCrowdsHead and PedVR / F2FCrowdsHead comparisons. Significant differences were observed for all the three comparisons for the Circle scene. Thus, the results prove the hypothesis 4 that a combination of the model of approach behavior and upper body motion generation is necessary to elicit a response from the users.

- **Question 5:** Question 5 evaluated whether the virtual agents seemed responsive. Friedman test revealed a significant difference in the perceived responsiveness of the virtual agents depending on the algorithm used for both the scenarios. For the Bidirectional scene, post hoc tests did not reveal a significant difference between the PedVR / F2FCrowds comparison, but significant difference was observed for the
Post hoc tests revealed significant differences for all the three comparisons for the Circle scene. Thus, the results prove the hypothesis 5 that the addition of model of approach behavior and upper body motion generation made the virtual agents appear more responsive.

- **Question 6:** We also asked the participants to report if they felt that the virtual agents responded even if the participant did not attempt to interact. Friedman test revealed a significant difference in the participant responses depending on the algorithm used for both the scenarios. Significant difference was not observed for the PedVR / F2FCrowds comparison in both the scenarios. Thus, the addition of the model of approach behavior does not make the virtual agents appear more responsive when the user does not attempt to interact. Significant differences were observed between the PedVR / F2FCrowdsHead comparisons for both scenarios and for F2FCrowds / F2FCrowdsHead comparison in the Bidirectional scenario. Thus, the combination of model of approach behavior and upper body motion generation makes the the virtual agents appear more responsive when the user does not attempt to interact. Some participants reported after the experiment that this made the virtual agents appear more “friendly” but further investigation is necessary.

7. Conclusion, Limitations, and Future Work

In this paper, we have presented techniques to compute the movements and trajectories of virtual agents to enable face-to-face interactions as part of a crowd. This includes an automatic approach for interaction velocity prediction, which we use to compute a collision-free velocity. We further augment the approach by simulating many upper body behaviors and movements. Our approach can simulate crowds with tens of agents at interactive rates, with support for F2F communications between the real user and virtual agents. We also performed a user study and concluded that our new algorithms increase the sense of social presence in
virtual environments. The virtual agents using our algorithms also appeared more responsive and were able to elicit more reaction from the users.

Our approach has some limitations. In particular, our criteria to trigger F2F interactions do not take into account the agent’s personality or emotions or the social norms. Furthermore, we only support a limited number of upper body movements or gestures. It would be useful to support verbal communication or conversations between the agents to increase the level of interaction. We would also like to model social signals like turn-taking and backchanneling, which are an important part of F2F interactions. We would like to evaluate our approaches in more complex scenarios, and compare with real-world scenarios. We use Oculus Rift to take user input. Since the walking area of Rift is limited, the users have to use a joystick or a keyboard which puts constraints on the realism of face-to-face interactions. We would like to use a wide area tracker framework to allow the real user to walk large distances in the physical world.

References


 REFERENCES


REFERENCES


REFERENCES


Nummenmaa, L., Hyönä, J., & Hietanen, J. K. (2009). I’ll walk this way: Eyes reveal the
direction of locomotion and make passersby look and go the other way. Psychological
Science, 20(12), 1454–1458.

approach for crowd simulation. In ACM Transactions on Graphics (TOG) (Vol. 29, 4,
p. 123). ACM.

Pantic, M., Cowie, R., D’Errico, F., Heylen, D., Mehu, M., Pelachaud, C., ... Vinciarelli, A.

high-density crowd simulation. In SCA ’07 (pp. 99–108).

Pelechano, N. [Nuria], Stocker, C., Allbeck, J., & 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-Volume 1 (pp. 136–
142).

Anxiety in Response to Three Different Types of Virtual Audience. Presence: Teleoper.
Virtual Environ. 11(1), 68–78. doi:10.1162/105474602317343668

Conference.

How to approach humans?-strategies for social robots to initiate interaction. In 2009


