

Watch Out! Modelling Pedestrians with Egocentric Distractions

Melissa Kremer
mkremer@cse.yorku.ca
York University
Toronto, Ontario, Canada

Mubbasir Kapadia
mk1353@cs.rutgers.edu
Rutgers University
New Brunswick, New Jersey, USA

Brandon Haworth
bhaworth@uvic.ca
University of Victoria
Victoria, British Columbia, Canada

Petros Faloutsos
pfal@eecs.yorku.ca
York University and UHN-Toronto Rehabilitation Institute
Toronto, Ontario, Canada

ABSTRACT

The use of mobile devices is one of the most commonly observed family of distracted behaviours exhibited by pedestrians in urban environments. We develop an event-driven behaviour tree model for distracted pedestrians that includes initiating mobile device use as well as terminating or pausing mobile device use based on internal or external cues to refocus attention. We present a simple, probabilistic attention model for such pedestrians. The proposed model is not meant to be complete. It primarily focuses on computing the probability that a distracted agent looks up, based on the agent's individual characteristics and the elements in their environment. We condition the potentially attention grabbing elements in the environment on distraction-specific egocentric fields for visual attention. We also propose an oriented ellipse model for capturing the affects of cognitively fuzzy goals during distracted navigation. Our model is simple and intuitively parameterized, and thus can be easily edited and extended.

CCS CONCEPTS

• **Computing methodologies** → **Agent / discrete models; Procedural animation.**

KEYWORDS

crowd simulation, distracted pedestrians, visual attention

ACM Reference Format:

Melissa Kremer, Brandon Haworth, Mubbasir Kapadia, and Petros Faloutsos. 2020. Watch Out! Modelling Pedestrians with Egocentric Distractions. In *Motion, Interaction and Games (MIG '20)*, October 16–18, 2020, Virtual Event, SC, USA. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3424636.3426910>

1 INTRODUCTION

The use of mobile devices, such as cell phones, while walking is a commonly observed behaviour among pedestrians in urban settings. Whether texting, talking or using other mobile apps the interaction between a pedestrian and a mobile device is often considered a

distraction, in the sense that the ability of the pedestrian to perceive and react to the environment is reduced. It is known that using a mobile device affects the distracted pedestrian's movement as well as focus, and can lead to serious accidents, such as tripping or walking into the path of a vehicle. At the same time it affects the nearby pedestrians who may be forced to slow down, or alter their trajectory. It is therefore timely and important for computer games and urban simulations, to model and account for distracted pedestrians.

This paper builds on previous work [Kremer et al. 2020] and presents a way of thinking about the problem of distraction modelling in synthetic crowds and a corresponding framework to build distracted agents. There are three key issues that one must address to incorporate distractions in a simulated environment: (a) when pedestrians decide to use their mobile devices, (b) how they behave when they are distracted, and (c) when they refocus their attention to their environment. Each one of these issues can be quite complex and is dependent on many variables, as they relate to cognition, vision, and attention. For instance, whether a pedestrian will answer or ignore a ringing cell phone, depends on the pedestrian's personality, age, location, whether they have company, and so many other circumstances. We propose a simple, editable, and extensible framework that one can use to author distracted pedestrians, and which offers common sense starting points, and intuitive parameters. Most of the important elements of this framework are modelled with conditional probabilities which can be set by the user based on intuition, experimentation, or derived from data and observations. In this paper, we exemplify the proposed framework by focusing on a few common distracted behaviours related to the use of mobile devices while walking: (a) taking a *phone call*, (b) *reading*, and (c) *texting*.

At the heart of our model is an agent-centric representation and abstraction of an agent's surrounding area, using egocentric fields that encode or condition probabilities that an agent's attention may be attracted by a specific area in his or her environment. Egocentric fields are a very effective and versatile way of encoding multiple layers of dynamic or static information centered on the agents or elements of the environment. A wide range of intuitive and powerful operations on these fields, such as filtering and convolution, can be used to combine and process them to support specific decision processes. An interesting example of the effectiveness of egocentric fields in crowd simulation is [Kapadia et al. 2009].

Our contributions can be summarized as follows:

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

MIG '20, October 16–18, 2020, Virtual Event, SC, USA

© 2020 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-8171-0/20/10.

<https://doi.org/10.1145/3424636.3426910>

- We blend the use of parametric and event-driven interactive behaviour trees to model state interactions between pedestrians and their devices.
- We ground the steering model parameters which are affected by different forms of distractions to the literature on human behaviour during distraction.
- We model navigation error (lateral deviation) identified in distraction studies by representing an agent's goal as a parameterized elliptical region.
- We model visual attention using Gaussian models in polar coordinates, and used them for both the steering and the attention refocusing models.
- We model the conditions for refocusing attention based on intrinsic parameters and extrinsic parameters conditioned on the visual attention model.

It is important to clarify that we do not propose a new steering method, but rather an egocentric behavioral layer that affects the steering abilities and parameters of an agent, and which can be combined with most of the standard steering approaches. In simple terms, this layer decides when and for how long a pedestrian temporarily acts more as a moving obstacle rather than a steering agent. The effect of the distracting behaviors on the sensory and locomotion abilities of the pedestrian is informed by the related literature which is based on actual observations, and thus reflects ground truth. The proposed intuitive parameterization can be easily adjusted as more studies become available in the future, or to fit a specific context.

2 RELATED WORKS

In this section we first present a short overview of crowd simulation techniques, and then we review the relevant literature in attention modelling and distracted pedestrians studies.

2.1 Crowd Simulation

A wide array of techniques have been used in crowd simulation. Many of these are outlined in [Thalmann and Musse 2012]. The Boids model proposed by Reynolds [Reynolds 1987] calculated velocities based on simple rules to perform agent navigation and obstacle avoidance, and was capable of demonstrating flock cohesion behaviour. More recently, anticipatory and synthetic vision based models have been proposed [Best et al. 2016; Ondřej et al. 2010; Paris et al. 2007; Van Den Berg et al. 2011]. Hybrid models have also been developed [Snape et al. 2011]. Predictive force-based models use current trajectories to compute future collision-free paths for some time window. One such model is Predictive Avoidance Model (PAM), which uses a piecewise predictive force to calculate future collisions well in advance so that the effort needed to make avoidance maneuvers can be minimized [Karamouzas et al. 2009]. Another method used a probabilistic model to improve the flexibility of predictions and the number of factors taken into account [Wolinski et al. 2016]. There is a lot of work on behavioural modeling for crowds, that focus on deliberate and complex behaviors such as those seen in a train station, for example [Shao and Terzopoulos 2005]. Most of this work is out of the scope of this paper, since we focus on the low level effects of egocentric and individual distractions.

In general, most of the prior work in crowd simulation focuses on normative agents with perfect visual abilities, and does not considered the effects of egocentric distractions, such as mobile devices, on the individual agent.

2.2 Gaze and Attention Modelling

A wide variety of approaches and techniques have been proposed for modelling vision and gaze in simulations, as examined in [Bruce et al. 2015]. This is indicative of how complicated the human vision system is and the difficulty of modelling it. VR-based studies have recently been proven to provide valuable insights into how people use gaze to navigate their environment [Berton et al. 2020, 2019; Huang and Terzopoulos 2020; Lynch et al. 2018]. One study showed that gaze is typically directed primarily towards neighbours that have the highest collision risk, based on the *distance to closest* approach and the *time to closest* approach [Meerhoff et al. 2018]. A similar study tracked users' eye gaze without a VR setup but using a monitor and a joystick [Berton et al. 2018].

Visual attention but without explicit gaze behaviour has been modelled to some degree within crowd simulations [Kuffner and Latombe 1999; Peters and O'Sullivan 2002]. Methods have been proposed that use environmental stimuli to trigger gaze redirection for characters based on a set of rules [Hill 1999; Khullar and Badler 2001].

Conversely, environment centric saliency maps have been utilized to draw agent gaze to the most noticeable objects [Itti et al. 1998]. In other works, gaze has been modelled in face-to-face scenarios for more personal situations [Bailly et al. 2010]. Gaze has also been modelled in interactive environments with real users [Kokkinara et al. 2011; Narang et al. 2016]. One approach [Grillon and Thalmann 2009] attempted to simulate gaze in crowds for aesthetics, although it didn't alter the trajectories of agents or investigate the effects of distraction on crowds. We are not aware of prior work in this area that models visual attention for distracted behaviours within our context.

2.3 Distracted Pedestrian Analysis

Previous studies on distraction in the real-world has shown various interesting effects of walking while distracted. For example, secondary activities can be visual, motor, or cognitive distractions or a combination of these. The type of distracting activity widely determines its task cost, with cognitive activities being the most distracting, followed by motor and then visual [Tian et al. 2018]. Another study suggests that gait is affected by cognitive and visual demand but not gross motor demand, although it did not examine or draw conclusions about the effects of fine motor demand such as swiping and tapping finger movements [Prupetkaew et al. 2019]. A study on driving while talking on the phone showed that participants talking on the phone recalled as little as half as much visual information they reported when not talking on the phone, and that hands-free phone calls were just as distracting as holding a cellphone while talking [Strayer et al. 2011].

3 METHODOLOGY

In this section, we discuss the details of our mobile device-centric distraction model. This includes models for initiating mobile device

use, the impact of distraction on steering, and reacting to internal and external cues to refocus. At the heart of our model are event-driven interactive behaviour trees which drive both agent distractions and mobile device events [Kapadia et al. 2015a,b]. That is, in this model, the objects of distraction (mobile devices) are separate autonomous agents which produce events that can be consumed at any time during simulation by pedestrians in possession of these devices. In addition to this behavioural level of modelling, we propose a ground-projected visual attention model. This model ties distraction types to changes in perception at the level of steering, such as changes in the Field of View (FOV) or neighbour distance. This model also serves as the basis for the proposed refocus approach, in which agents can become attentive again because of their local surroundings.

3.1 Distractions

We implement three types of distraction related to mobile devices, which are prevalent in everyday modern lives, particularly in urban settings. These include taking a *phone call*, *reading*, and *texting* on a mobile device while navigating.

At a high level, these behaviours are managed and triggered through two types of behaviour trees. A mobile device agent has a parametric behaviour tree which drives the use of the device. This behaviour tree triggers events based on distraction specific probabilities set by the animator. These are p_c , p_r , and p_t for *phone call*, *reading*, and *texting* respectively, such that $p_c + p_r + p_t + p_n = 1$. Here $p_n = 1 - (p_c + p_r + p_t)$ is simply the probability that no mobile device interaction will begin during a time-step. At each simulation time-step $t \in T$, a trial in a multinomial distribution is performed, such that the number of trials is $n = T f_s$, where f_s is the frequency of simulation updates. In this way, an animator may choose values for p_c , p_r , and p_t such that a pedestrian's agent will likely become distracted over some period. Once a particular distraction trial succeeds, a distraction specific event is triggered.

The distraction event is consumed by the pedestrian agent's event driven interactive behaviour tree. This causes a number of distraction specific changes in the pedestrian agent's animation, steering, and visual attention model. In our proposed approach, animation is driven by a typical Inverse Kinematics (IK) approach with procedural end-effector targets that include noise and variation among the pedestrian agents to improve naturalness and heterogeneity of large scenarios. The impact on steering and attention are explained in the following sections.

3.2 Steering

The proposed approach includes several distraction specific changes to steering that are based on the literature on human navigation during distractions. The primary parameters related to steering are preferred speed, waypoint undershoot/overshoot, waypoint lateral deviation, field of view (FOV), and neighbour distance. The FOV and neighbour distance parameters are detailed in Section 3.3 as they relate to the visual attention model. In our implementation we use PAM [Karamouzas et al. 2009] as the underlying steering model. However, any steering model may be used as long as it uses the aforementioned parameters and a local goal waypoint as its current target.

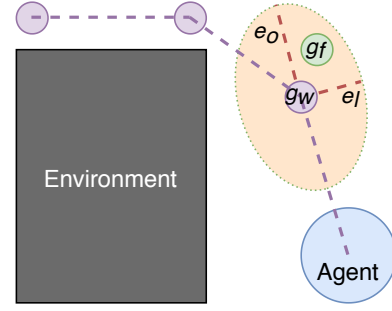


Figure 1: A birds-eye view illustration of the proposed fuzzy goal model in a simple environment. We encode the possibility of lateral deviation and overshoot/undershoot as the lengths of the semi-minor e_l and semi-major e_o axes respectively. This forms an oriented ellipse around the current waypoint goal g_w within which the new fuzzy goal g_f is randomly chosen.

The preferred speed s_p for a normative, not distracted, pedestrian agent can vary depending on the context of the scenario. For illustration purposes we use a default value of 1.3m/s, which reflects the average human walking speed. The preferred walking speed during distraction may be a fraction of the agent's normative walking speed depending on the specifics of the distraction model. For the implementation of mobile device-based distractions in the proposed approach, we derive values from a review of the literature on distractions. For a pedestrian agent on a *phone call*, the preferred speed is set to 1.066 m/s or 82% of the normative walking speed [Prupetkaew et al. 2019]. For a pedestrian agent *reading*, the preferred speed is set to 1.144 m/s (88%) [Niederer et al. 2018; Schabrun et al. 2014]. For a distracted agent *texting*, the preferred speed is set to 1.04 m/s or (80%) [Agostini et al. 2015; Cha et al. 2015; Haga et al. 2015; Licence et al. 2015; Pizzamiglio et al. 2017; Plummer et al. 2015; Prupetkaew et al. 2019; Schabrun et al. 2014; Yu et al. 2015]. Further research may be needed to determine distributions for values for these parameters as they relate to age, sex, culture, etc.

During *reading* or *texting* behaviours, where vision is primarily focused downward and the visual horizon is short, navigating pedestrians must rely on a mental model of their local waypoint. To model this fuzzy mental model of local waypoints we propose a parametric ellipse model of the local waypoint. This model encapsulates both lateral deviation and overshoot/undershoot caused by a less than accurate cognition of local waypoints. Lateral deviation means that someone who is distracted begins to deviate left or right from a straight path to their local waypoint, and it has been extensively studied in the related literature [Jeon et al. 2016; Lamberg and Muratori 2012; Schabrun et al. 2014]. Overshoot/undershoot is an observed real-world behaviour of pedestrians who are so distracted that they forget to turn where they should or proceed past their destination. Both of these behaviours are modelled by sampling the proposed oriented parametric ellipse, as seen in Figure 1. The semi-major, e_o , and semi-minor, e_l axes govern the degree of undershoot/overshoot and lateral deviation respectively. In practice,

the semi-major and semi-minor axes of the ellipse are determined by the distance of the agent to their actual waypoint, g_w , when the fuzzy waypoint g_f is chosen. However, an animator may specify a maximum length of the semi-major and semi-minor axes of the ellipse per agent, and at runtime the model will attempt to keep this ratio intact. To achieve this, the semi-minor axis is set to 10% of the distance between the agent and their current waypoint, up to the specified maximum. The semi-major axis length is set to 10% of this distance multiplied by the major-minor axis ratio derived from the user specified maximums of the ellipsoid axes. We use the 10% value in our experiments because it corresponds to the average largest lateral deviation amount reported in the literature [Jeon et al. 2016; Lamberg and Muratori 2012; Schabrun et al. 2014]. This adaptive approach also ensures that the generated fuzzy waypoint is never behind the agent. The ellipse representation is then sampled for a random point within its bounds, which we refer to as the fuzzy waypoint, g_f . Once a potential fuzzy waypoint is chosen, it is tested and adapted for reachability against the underlying environment model (Navigation Mesh, Grid, Obstacles, etc). If an agent refocuses, i.e. is no longer distracted, their local waypoint returns to their original waypoint. If an agent reaches a fuzzy waypoint while distracted, the true waypoint is updated as normal. If the agent is still distracted, a new fuzzy waypoint is then computed for them, and so on.

3.3 Visual Attention

We propose a vision model for normative and distracted pedestrians that is designed to interface with both the underlying steering model and the proposed model for refocusing attention. Visual attention is a complex process which incorporates intrinsic and extrinsic factors into visual perception [Tsotsos 2001]. To overcome this complexity, and for the sake of usability and performance, we make use of the fact that human visual perception clarity is concentrated in the centre of the visual field and becomes less focused towards the edges of the visual field. Additionally, people tend to focus more attention on visual cues central to their visual field and more nearby to themselves [Ikeda and Takeuchi 1975; Leibowitz and Appelle 1969]. In our model, we collapse the three-dimensional field of view formed by binocular vision into a flat model and consider horizontal vision, similar to how steering models perceive environment. The proposed model represents the visual field in polar coordinates over a 200° sweep. Within this visual field we represent the distribution of attention as two Gaussians, one half Gaussian in the r dimension and one full Gaussian in the θ dimension of the polar field. These represent the fall off of attention with respect to distance and the attenuation of attention with respect to angle (i.e. peripheral vision) respectively. The attenuation with respect to distance is defined as $\mathcal{N}(0, \sigma_r)$, where $r = 0$ is the centre of the head. The attenuation of the attention with respect to angle is defined as $\mathcal{N}(0, \sigma_\theta)$, where $\theta = 0$ is aligned with the forward vector of the agent's head. In this way, each visual attention model for each distraction has two parameters (σ_r, σ_θ). These functions are convolved together to produce a single distribution of attention over the polar visual field. The relative values of the resultant field can then be transformed and stored in a single channel of a texture for fast lookup. In this paper, we normalize individual values in

the field textures and sample singular points of the field, using the nearest point on object's (agent or obstacle) underlying collision representation. In this way, each sample of each individual pixel in the texture can be modelled by a Bernoulli distribution conditioned on its location in the field $p(x|r, \theta)$. However, this method affords many approaches, such as integration over a probability density function (by normalizing pixel values by the sum of pixels in the field, then summing the pixels within the region of overlap with the field and the object). As well, a practitioner may choose different parameters or functions to convolve over (r, θ) . The construction of each visual attention model presented in this paper is explained in the following paragraphs and can be seen in Figure 2.

A pedestrian with normative vision during walking is often modelled with a neighbour distance [Karamouzas et al. 2009], planning horizon [Singh et al. 2011], or query radius [Helbing and Molnar 1995] between 5 – 15m. This range of choices are backed up by literature on human-human collision interactions with a bias toward lower values [Cinelli and Patla 2008; Fajen and Warren 2003]. In this work, to support arbitrary models, our visual field is constrained to 10m or the average query value for neighbouring objects used in steering models. The two parameters for normative vision of a pedestrian agent are then $\sigma_r = 5m$ and $\sigma_\theta = 30^\circ$, to utilize the entirety of the field where the majority of attention is within 5m and within a 60° sweep, which corresponds to the reported central visual field horizontal range [Bhise 2011].

We assume that an agent talking on the phone is able to see as far as an undistracted agent, so $\sigma_r = 5m$. However, studies have shown that while talking on the phone, visual information is often missed and peripheral vision shrinks about 7 – 10% for large or small targets due to the cognitive load [Maples et al. 2008]. Considering other agents as large targets, we shrink σ_θ by about 7% and set $\sigma_\theta = 28^\circ$. This means that the majority of the visual field is still used but attention is focused within 5m and a 56° sweep.

A reading agent is a more complicated case because the head is tilted downwards so vision is significantly affected. A study on peripheral vision showed that when people focus on visual information, their ability to notice and process information in their peripheral vision diminishes [Ikeda and Takeuchi 1975]. Further, the study shows that the peripheral information loss is correlated with the foveal load. However, they also found that the effect is reduced if the subjects have participated in the study before and are thus trained to notice things in their peripheral vision. Trained subjects had an average peripheral information uptake of about 76.5% for high foveal load when compared to normative participants with no visual load. For our experiments, we assume that cellphone users frequently use their cellphones while walking and are therefore trained to use their peripheral vision while walking, and we also assume that cellphone use is a high foveal load. Therefore we set σ_θ to 76.5% of normative, so $\sigma_\theta = 22.95^\circ$ for reading agents. Another study [Schabrun et al. 2014] reports that people reading on a cell phone have an average head pitch angle of 29.22° downward. Reported values for the average height for adults age 18 years in 1996 was 1.71m for men and 1.59m for women worldwide so we used an average of these two values, 1.65m, to calculate a focus distance of $\sigma_r = 2.95m$ on the ground in front of them [Collaboration et al. 2016]. Therefore, attention is focused within 2.95m in front of them and within a 46° sweep.

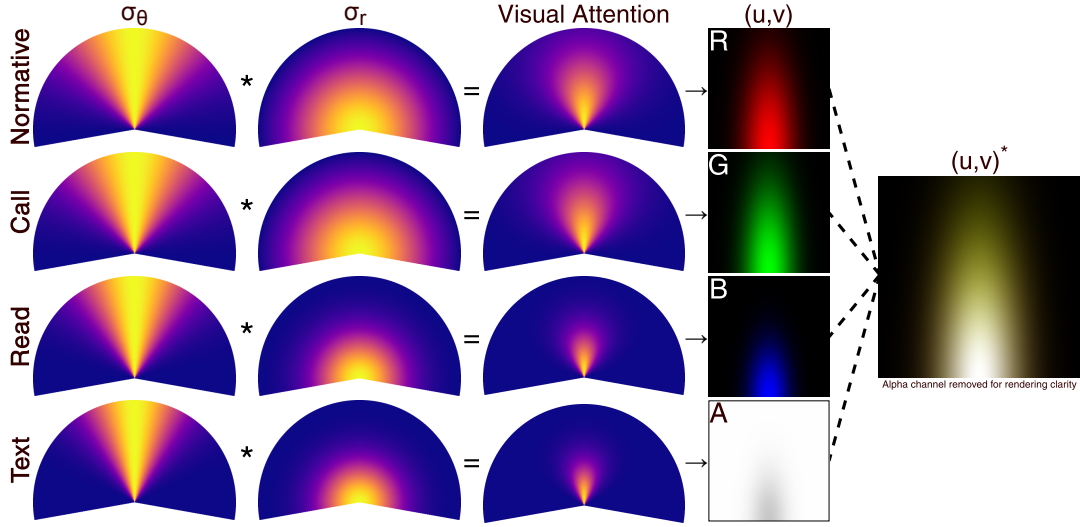


Figure 2: The four visual attention models, combined into an RGBA texture.

Similar to a reading agent, a texting agent also has reduced field of view due to looking down. However, texting also has an additional motor load and is perhaps more cognitively demanding than reading. We therefore estimate an additional 10% loss in peripheral vision information processing and set σ_θ to 66.5% of normative, so $\sigma_\theta = 19.95^\circ$ for texting agents. However, [Schabrun et al. 2014] reports that texting pedestrians pitch their head downwards by a larger amount than reading pedestrians, about 31.80° which we can again use with an average height of $1.65m$ to set $\sigma_r = 2.66m$. Attention is then focused within $2.66m$ and a 40° sweep.

A visual representation of each model and how its respective texture is created is shown in Figure 2. In the first three columns, bright yellow areas represent the space where most attention is paid, while no attention is paid to dark blue areas.

3.4 Refocusing

We propose a refocusing model which produces one of two actions *refocus*, *continue*. For a pedestrian agent which is currently distracted, a reaction module executes three possible Bernoulli trials at a frequency f_r . We use the value $f_r = 3Hz$, an approximation of the inverse of average human reaction time. However, this may be set to a lower or higher value per agent, producing less or more attentive pedestrians respectively.

The first trial represents the intrinsic properties of a pedestrian and tests whether the agent is going to refocus because of his or her circumstance and personality. This trial is conducted at every refocus update. The value of $Pr(refocus) = p_i$ may represent whether a pedestrian agent is a person who takes important calls (e.g. a business person, on-call worker, etc.), or other elements of the agent's personality. This value may be set by the practitioner to suit scenario needs. As this value remains fixed, at every reaction update the Bernoulli trial is a part of a Bernoulli process which can be modelled as a Binomial distribution $\mathcal{B}(N_r, p_i)$ where N_r is the total number of reaction updates $N_r = T_d f_r$ and T_d is the duration of the distraction up to the timestep $t \in T$.

The second and third trials represent the extrinsic properties of local objects (obstacles, agents) relative to the pedestrian agent. Both of these trials are conditioned on the visual attention model and each is performed for each object in the FOV at every refocus update. The first of these is related to the static visual appeal of the object and tests whether the agent will notice the object and refocus his or her attention. The value of $Pr(refocus) = p_e$ may represent shape, colour, contrast, importance, and so on of an object. The probability an agent will see the object because of its properties is conditioned on the visual attention field such that $P(refocus|p_e, r, \theta)$. The third trial accounts for an object's relative speed which is known to have particular importance for predicting collisions [Karamouzas et al. 2014], and thus we consider it separately in our context. The value

$$Pr(refocus) = p_v = \frac{|v_{rel}|}{2|v_{max}|} \quad (1)$$

represents whether an agent will refocus because of the movement of an object, where v_{max} is the maximum velocity of all the objects in the scene. The probability that an agent will see the object because of its movement is also conditioned on the visual attention field such that $P(refocus|p_v, r, \theta)$.

When a distracted activity is chosen, a random task time is selected. The maximum task time is parameterizable and can be specified by the practitioner. If an agent is interrupted while distracted and looks up, after another random wait time the agent will resume the task, provided that there is at least 1 second of task time remaining, to avoid oscillations.

When one of the trials succeeds then the associated agent switches to a normative visual attention model. In the reading and texting cases the agent also interrupts temporarily the distracting activity and looks up and lowers the device, while in the talking case the agent who is already looking up continues to hold the device to a talking position.

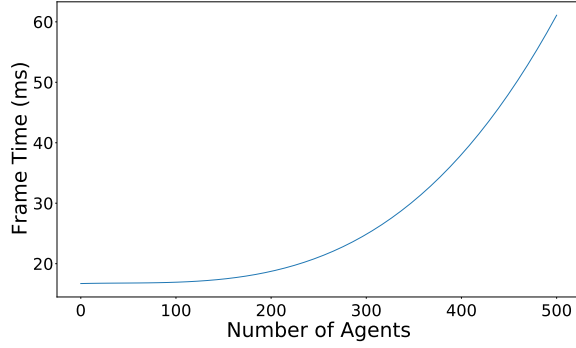


Figure 3: The frame time in milliseconds as the number of agents increases.

3.5 Crowd Authoring

In order to test our method with virtual crowds quickly and easily, we have implemented a simple and intuitive crowd authoring interface in Unity3D. For every test scene we set one or more goals around the scene, and crowd spawning areas. Crowd spawning areas are simple planes placed in the scene that represent the dimensions or areas within which agents should be spawned. The user drags the desired crowd spawning area and goal from the scene hierarchy into the floor and goal fields of the window respectively. A variety of options are then available to the user, including how many agents of each type to generate. Normative agents do not have cellphones and will not become distracted, while distracted agents may occasionally become distracted or undistracted according to their properties and the corresponding visual attention model. Opening up the distraction parameters foldout allows the user to quickly set desired probabilities and other parameters for distracted agents all in one place. Clicking the Create Agents button will spawn the desired number of agents with the desired parameters into the scene view and hierarchy, where their positions and other attributes can be further tweaked. The user is free to spawn multiple groups of agents with different parameters in the same or different spaces and with the same or different goals. A similar runtime tool has also been implemented that determines agent positions randomly at every runtime.

When agents are generated, a free location is found by sampling the navmesh randomly within the spawn region. When a free location is found with appropriate distance to other agents, the agent is spawned. Obstacles can also be spawned with or without agents, and the navmesh is updated accordingly. An equally simple interface is used to set the parameters of the objects, such as the extrinsic probability, p_e , that controls an object's visual appeal.

4 RESULTS

4.1 Computational Performance

In this section we study the computational performance of our framework using a variant of the statistical approach proposed in [Kapadia et al. 2011]. At each trial, an environment is populated with 100 obstacles randomly placed in a grid like fashion, and then agents and their goals are randomly generated.

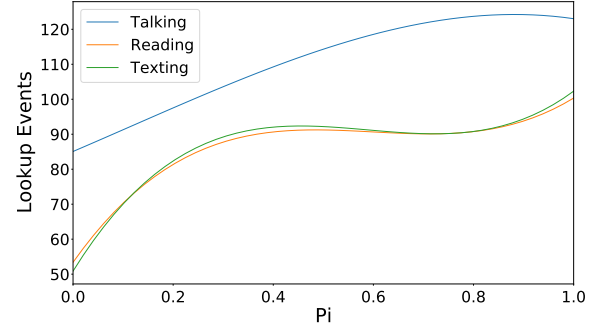


Figure 4: The number of lookups across distraction types for increasing p_i values.

We use this approach with an increasing number of distracted agents and measure the frame time in milliseconds. The result is illustrated in Figure 3. The system is capable of handling more than 200 distracted agents without a significant increase in frame time. However, the frame time starts to increase as the number of agents increases towards 500. It is worth noting though that our implementation has not been optimized yet.

4.2 Refocus Model Parameters

In order to demonstrate the effect of the intrinsic look up probability p_i , we use the scenario space environment with a fixed number of 20 agents. The p_i value is homogeneous among all agents, but we slowly increase this value from 0 to 1 and count the number of refocus events. The result is shown in Figure 4. As expected, higher values of p_i yields a higher number of refocus events. Interestingly, the talking distraction exhibits more look up behaviours than the other distracted activities. This makes sense, because the visual field of these agents is much larger so it is more likely that objects in the environment represented in this process by the extrinsic probabilities p_e and p_v will be noticed and trigger a look up behaviour.

The same procedure is used to demonstrate the effect of the extrinsic look up probability p_e . For a group of homogeneous agents, we use the same p_e factor for all obstacles and slowly increase it from 0 to 1, and count the number of refocus events. The result is shown in Figure 5. Higher values of p_e correlate to higher numbers of refocus events, as one would expect.

The impact on look up count plateaus after p_i and p_e values have reached certain points. Since the probability checks run very frequently at three times per second, even small increases for p_i and p_e cause a look up behaviour to trigger, up to a point where further increasing the value has only a small impact.

Finally, we evaluate the effect of the p_v by comparing unidirectional and bidirectional flows. In a unidirectional flow scenario, relative velocity between agents will be small since all agents are moving in the same direction. In a bidirectional flow, agents approaching from the opposite direction will have high relative velocities. The number of refocus events is shown in Figure 6.

As expected, talking pedestrians exhibit a higher number of look up behaviours. Surprisingly, however, texting agents had slightly

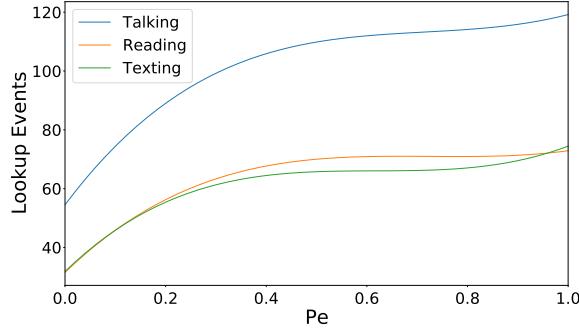


Figure 5: The number of lookups across distraction types for increasing p_e values.

more look up behaviours than reading agents. A possible explanation for this is that texting agents move on average slower than reading agents (see Section 3.2), so external objects and agents may be in their field of view for longer, resulting in a higher chance of catching their attention and triggering a look up behaviour. Also, because texting agents move slower they remain in the scene longer, which allows them more opportunity for look up triggers.

4.3 Visual Attention Model Parameters

To illustrate and compare results from the different vision models for each of the distraction types, we again use 20 agents in the scenario space environment. All parameters are homogeneous across agents, however in each trial we restrict the attention models so that only one model can be active, either normative, reading on a device, talking, or texting. We then record the refocus events, collisions, and average path length. Collisions are detected through Unity3D’s built-in collision detection where each agent is represented by a capsule collider. Thus we can compare how these measurements differ between the visual attention models for each activity type. The results are summarized in Figure 7.

As expected, normative and talking modalities had similar results for collisions, while texting and reading distractions resulted in more collisions. Unexpectedly, reading had the most collisions. Even though the visual field is slightly more reduced for texters than readers, this result could be explained by the fact that readers also move quicker than texters resulting in more collisions. For look up events, talking had the highest amount while texting and reading were roughly the same. This is expected, as for talkers looking up does not interrupt the distraction, so they tend to be distracted for longer and so would trigger more look up events. Path length was highest for readers and texters, with readers’ paths being longer than texters. This may be explained by the increased collision rate for readers, which causes them to be pushed away from the optimal path.

4.4 Fuzzy Goal Model Parameters

To evaluate the parameters used in the fuzzy goal model, we use the scenario space environment with a single agent. For this agent, random values for the maximum semi-major and semi-minor lengths of the fuzzy goal ellipse are generated along with a randomized

goal. Here we randomize the maximum semi-major and semi-minor axis between $[0, 1]$, but higher values are also possible. Then we measure lateral deviation and overshoot/undershoot with respect to the actual waypoint. We also measure the pathlength. The results are shown in Figure 8.

From this we can see that in general path length is higher when the maximum semi-minor axis length is high. This is expected, since semi-minor axis length represents the amount of lateral deviation, which deviates from optimal paths. Generally higher values of semi-major axis length also increase path length, but to a lesser extent and with more noise. Overshooting or undershooting may not result in significant deviations from the optimal path. In particular, undershooting is not likely to affect path length as much because upon the agents reaching their fuzzy goal they will start heading to the next waypoint which is often in the same direction.

4.5 Qualitative Evaluation

We evaluate qualitatively the proposed framework using the following common and indicative scenarios: (a) two oncoming agents with obstacles, and (b) two groups of agents crossing a corridor from opposite sides.

In the oncoming obstacle scenario, two agents start at opposite sides and move along potentially crossings paths at the presence of obstacles. Figure 9 shows the resulting trajectories for the three different distractions with both agents performing the same distracted activity. Normative and talking models exhibit similar trajectories, while the reading agents show deviations from the optimal path. The trajectory of the texting agent that started on the left side, clearly shows that the agent bumped on the first obstacle that the agent encountered.

In the crossing groups scenario, two groups of a total of 50 agents walk through a narrow corridor from opposite directions. The results for this scenario are depicted in Figure 10. For the texting and reading case, distraction tends to occur less in the middle where the two groups meet, because the p_v factor is triggering look up behaviours, while the talking on the phone distraction is not interrupted. Figure 11 shows the traces for the same scenario with twenty agents distracted in all three different ways.

In Figure 12, we show the visual field of a single distracted agent from this scenario before and after a look up behaviour triggers due to the presence of another agent, for reference.

In Figure 13 we show a snapshot from an urban scene where pedestrians, including distracted ones, cross at an intersection. We refer the reader to the accompanying video for animated versions of these experiments which show the resulting behaviours more clearly.

5 DISCUSSION

The proposed framework for authoring distracted pedestrians and the associated visual attention model are specifically demonstrated on egocentric distractions related to the use of mobile devices. A number of intuitive parameters are available to the user to author specific situations, personalities, and variations. For example, very cautious agents may not become distracted as often or may look up more often. Outgoing or less cautious agents may send and receive more texts and calls and so on. In particular, the attention model,

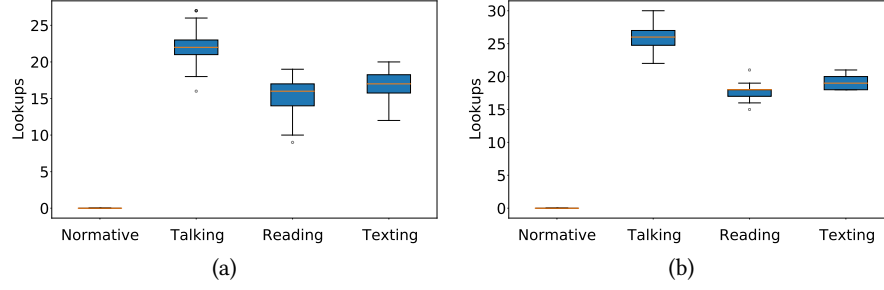


Figure 6: The number of lookups across distraction models in (a) uni-directional and (b) bi-directional flows.

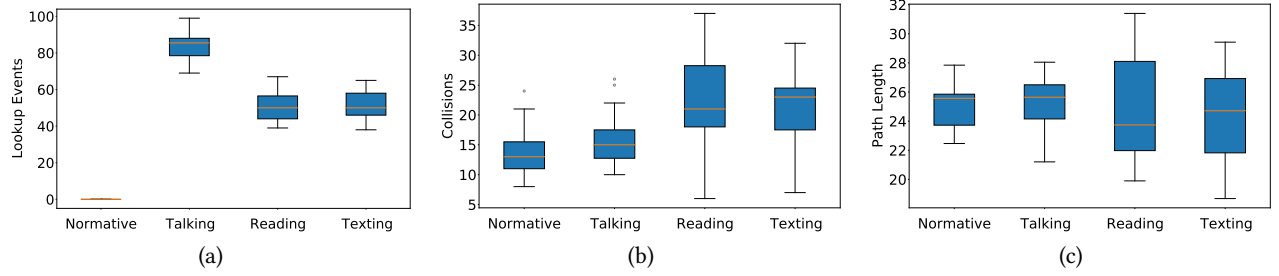


Figure 7: Quantitative comparison across distractions based for (a) look up events, (b) collisions, and (c) path length.

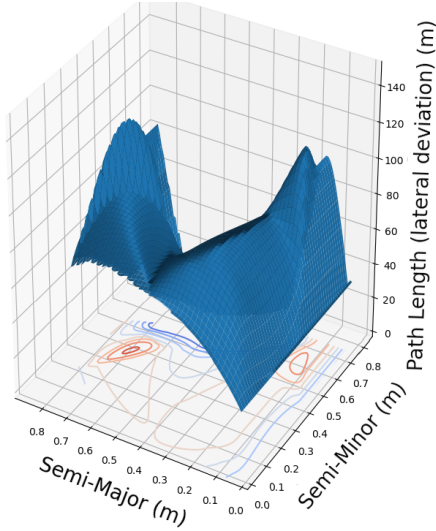


Figure 8: The average agent path length as a function of the semi major and semi minor fuzzy goal ellipse parameters.

which is based on egocentric fields and related operations, can be customized or extended to support agents with different sensory abilities, such as limited peripheral vision and other vision impairments. Furthermore, it could be combined with more sophisticated models of agent vision that have been reported in the computer graphics and computer vision literature.

Our work focuses on individual (egocentric) rather than emerging group behaviours. Nevertheless, comparing the resulting traces

with real-world crowd data may offer additional insights. We plan to address this issue in the future with a carefully designed study.

In the future, we aim to address non-egocentric distractions, and deliberate behaviours that may impede an agent's navigation abilities, such as window shopping, reading an advertisement on a board, looking at an interesting object or an accident, talking to a friend, et al. It would be interesting to investigate whether projecting variants of the proposed attention models along the direction of gaze would be sufficient to accommodate such behaviours, rather than resorting to a 3D representation of the field of view.

Exploring the range of the distracted agents' parameters, as well as determining and providing default values from real data are important future tasks. We also plan to investigate how other individual characteristics of an agent, such as age, gender, mobility and sensory abilities, may affect the agent's behaviour towards egocentric distractions.

6 CONCLUSION

We have presented an approach and a framework for modeling egocentric distractions related to the use of mobile devices for crowds in everyday casual navigation tasks. Our approach provides an intuitive set of parameters that can be used to customize these behaviors for each agent. The parameters of the model and their default values are grounded in studies of distracted pedestrians, and thus reflect ground truth.

The proposed attention model can be easily extended to model pedestrians with varying sensory and physical abilities. We believe that the inclusion of a wide variety of non-normative pedestrian behaviours and attributes will push realism and fidelity in crowd animations to a higher level that more accurately reflects the diversity we see in our daily lives.

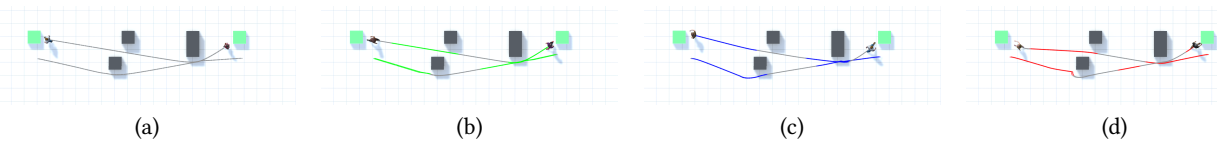


Figure 9: The paths taken in the oncoming obstacles scenario by (a) two normative agents, (b) two distracted agents talking on the phone, (c) two distracted agents reading, and (d) two distracted agents texting. Green, blue, and red highlighted trajectories indicate where along their path an agent was distracted.

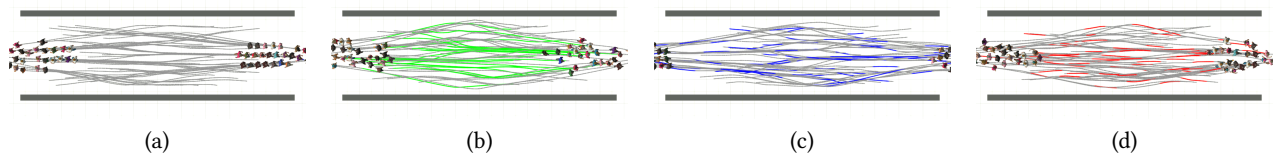


Figure 10: The paths taken in the crossing groups scenario by (a) 50 normative agents, (b) 30 normative and 20 distracted agents talking on the phone, (c) 30 normative and 20 distracted agents reading, and (d) 30 normative and 20 distracted agents texting. Green, blue, and red highlighted trajectories indicate where along their path an agent was distracted.

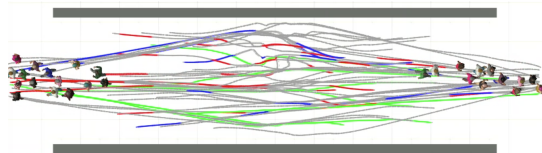


Figure 11: The paths taken in the crossing group scenario with 30 normative and 20 distracted agents with all three different types of distraction.

ACKNOWLEDGMENTS

Funding was provided by Ontario Research Foundation (Grant No. RE08-054) and National Science Foundation (Grant Nos. IIS1703883, S&AS-1723869).

REFERENCES

- Valentina Agostini, Francesco Lo Fermo, Giuseppe Massazza, and Marco Knaflitz. 2015. Does texting while walking really affect gait in young adults? *Journal of neuroengineering and rehabilitation* 12, 1 (2015), 86.
- G  rard Bailly, Stephan Raidt, and Fr  d  ric Elisei. 2010. Gaze, conversational agents and face-to-face communication. *Speech Communication* 52, 6 (2010), 598–612.
- Florian Berton, Ludovic Hoyet, Anne-H  l  ne Olivier, Julien Bruneau, Olivier Le Meur, and Julien Pettr  . 2020. Eye-Gaze Activity in Crowds: Impact of Virtual Reality and Density. In *IEEE Virtual Reality Conference*.
- Florian Berton, Ludovic Hoyet, Anne-H  l  ne Olivier, and Julien Pettr  . 2018. Gaze Behaviour During Collision Avoidance Between Walkers: A Preliminary Study to Design an Experimental Platform.
- Florian Berton, Anne-H  l  ne Olivier, Julien Bruneau, Ludovic Hoyet, and Julien Pettr  . 2019. Studying Gaze Behaviour During Collision Avoidance With a Virtual Walker: Influence of the Virtual Reality Setup. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 717–725.
- Andrew Best, Sahil Narang, and Dinesh Manocha. 2016. Real-time reciprocal collision avoidance with elliptical agents. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 298–305.
- Vivek D Bhise. 2011. *Ergonomics in the automotive design process*. CRC Press.
- Neil DB Bruce, Calden Wloka, Nick Frosst, Shafin Rahman, and John K Tsotsos. 2015. On computational modeling of visual saliency: Examining what’s right, and what’s left. *Vision research* 116 (2015), 95–112.
- Jaeyun Cha, Hyunjin Kim, Jaemyoung Park, and Changho Song. 2015. Effects of mobile texting and gaming on gait with obstructions under different illumination levels. *Physical Therapy Rehabilitation Science* 4, 1 (2015), 32–37.
- Michael E Cinelli and Aftab E Patla. 2008. Locomotor avoidance behaviours during a visually guided task involving an approaching object. *Gait & posture* 28, 4 (2008), 596–601.
- NCD Risk Factor Collaboration et al. 2016. A century of trends in adult human height. *Elife* 5 (2016), e13410.
- Brett R Fajen and William H Warren. 2003. Behavioral dynamics of steering, obstacle avoidance, and route selection. *Journal of Experimental Psychology: Human Perception and Performance* 29, 2 (2003), 343.
- Helena Grillon and Daniel Thalmann. 2009. Simulating gaze attention behaviors for crowds. *Computer Animation and Virtual Worlds* 20, 2-3 (2009), 111–119.
- Shigeru Haga, Ayaka Sano, Yuri Sekine, Hideka Sato, Saki Yamaguchi, and Kosuke Masuda. 2015. Effects of using a smart phone on pedestrians’ attention and walking. *Procedia Manufacturing* 3 (2015), 2574–2580.
- Dirk Helbing and Peter Molnar. 1995. Social force model for pedestrian dynamics. *Physical review E* 51, 5 (1995), 4282.
- Randall Hill. 1999. Modeling perceptual attention in virtual humans. In *Proceedings of the 8th Conference on Computer Generated Forces and Behavioral Representation*. Citeseer, 563–573.
- W. Huang and D. Terzopoulos. 2020. Door and Doorway Etiquette for Virtual Humans. *IEEE Transactions on Visualization and Computer Graphics* 26, 3 (2020), 1502–1517.
- Mitsuo Ikeda and Tetsuji Takeuchi. 1975. Influence of foveal load on the functional visual field. *Perception & Psychophysics* 18, 4 (1975), 255–260.
- Laurent Itti, Christof Koch, and Ernst Niebur. 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 11 (1998), 1254–1259.
- SoYeong Jeon, ChoRong Kim, SunHae Song, and GyuChang Lee. 2016. Changes in gait pattern during multitask using smartphones. *Work* 53, 2 (2016), 241–247.
- Mubbasir Kapadia, Jessica Falk, Fabio Z  nd, Marcel Marti, Robert W Sumner, and Markus Gross. 2015a. Computer-assisted authoring of interactive narratives. In *Proceedings of the 19th Symposium on Interactive 3D Graphics and Games*. 85–92.
- Mubbasir Kapadia, Shawn Singh, William Hewlett, and Petros Faloutsos. 2009. Egocentric Affordance Fields in Pedestrian Steering. In *Proceedings of the 2009 Symposium on Interactive 3D Graphics and Games* (Boston, Massachusetts). Association for Computing Machinery, New York, NY, USA, 215–223. <https://doi.org/10.1145/1507149.1507185>
- Mubbasir Kapadia, Matt Wang, Shawn Singh, Glenn Reinman, and Petros Faloutsos. 2011. Scenario Space: Characterizing Coverage, Quality, and Failure of Steering Algorithms. In *Proceedings of the 2011 ACM SIGGRAPH/EG Symposium on Computer Animation* (Vancouver, British Columbia, Canada). ACM, New York, NY, USA, 53–62. <https://doi.org/10.1145/2019406.2019414>
- Mubbasir Kapadia, Fabio Z  nd, Jessica Falk, Marcel Marti, Robert W Sumner, and Markus Gross. 2015b. Evaluating the authoring complexity of interactive narratives with interactive behaviour trees. *Foundations of Digital Games* (2015).
- Ioannis Karamouzas, Peter Heil, Pascal Van Beek, and Mark H Overmars. 2009. A predictive collision avoidance model for pedestrian simulation. In *International workshop on motion in games*. Springer, 41–52.
- Ioannis Karamouzas, Brian Skinner, and Stephen J Guy. 2014. Universal power law governing pedestrian interactions. *Physical review letters* 113, 23 (2014), 238701.

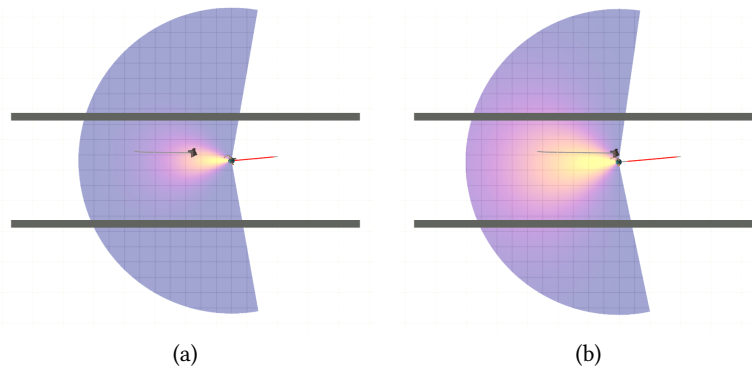


Figure 12: The visual field of a distracted texting agent (a) before and (b) after a look up behaviour triggers.



Figure 13: A snapshot from an intersection crossing that includes distracted pedestrians.

- Sonu Chopra Khullar and Norman I Badler. 2001. Where to look? Automating attending behaviors of virtual human characters. *Autonomous Agents and Multi-Agent Systems* 4, 1-2 (2001), 9–23.
- Elena Kokkinara, Oyewole Oyekoya, and Anthony Steed. 2011. Modelling selective visual attention for autonomous virtual characters. *Computer Animation and Virtual Worlds* 22, 4 (2011), 361–369.
- Melissa Kremer, Brandon Haworth, Mubbasir Kapadia, and Petros Faloutsos. 2020. Modelling distracted agents in crowd simulations. *The Visual Computer* (2020), 1–12.
- James J Kuffner and J-C Latombe. 1999. Fast synthetic vision, memory, and learning models for virtual humans. In *Proceedings Computer Animation 1999*. IEEE, 118–127.
- Eric M Lamberg and Lisa M Muratori. 2012. Cell phones change the way we walk. *Gait & posture* 35, 4 (2012), 688–690.
- HW Leibowitz and Stuart Appelle. 1969. The effect of a central task on luminance thresholds for peripherally presented stimuli. *Human Factors* 11, 4 (1969), 387–391.
- Sammy Licence, Robynne Smith, Miranda P McGuigan, and Conrad P Earnest. 2015. Gait pattern alterations during walking, texting and walking and texting during cognitively distractive tasks while negotiating common pedestrian obstacles. *PLoS one* 10, 7 (2015), e0133281.
- Sean D Lynch, Julien Pettré, Julien Bruneau, Richard Kulpa, Arnel Crétual, and Anne-Hélène Olivier. 2018. Effect of virtual human gaze behaviour during an orthogonal collision avoidance walking task. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 136–142.
- WC Maples, Wes DeRosier, Richard Hoes, Rodney Bendure, and Sherl Moore. 2008. The effects of cell phone use on peripheral vision. *Optometry-Journal of the American Optometric Association* 79, 1 (2008), 36–42.
- LA Meerhoff, Julien Bruneau, A Vu, A-H Olivier, and Julien Pettré. 2018. Guided by gaze: Prioritization strategy when navigating through a virtual crowd can be assessed through gaze activity. *Acta psychologica* 190 (2018), 248–257.
- Sahil Narang, Andrew Best, Tanmay Randhavan, Ari Shapiro, and Dinesh Manocha. 2016. PedVR: Simulating gaze-based interactions between a real user and virtual crowds. In *Proceedings of the 22nd ACM conference on virtual reality software and technology*. ACM, 91–100.
- Daniel Niederer, Anke Bumann, Yvonne Mühlhauser, Mareike Schmitt, Katja Wess, Tobias Engeroff, Jan Wilke, Lutz Vogt, and Winfried Banzer. 2018. Specific smartphone usage and cognitive performance affect gait characteristics during free-living and treadmill walking. *Gait & posture* 62 (2018), 415–421.
- Jan Ondřej, Julien Pettré, Anne-Hélène Olivier, and Stéphane Donikian. 2010. A synthetic-vision based steering approach for crowd simulation. *ACM Transactions on Graphics (TOG)* 29, 4 (2010), 1–9.
- Sébastien Paris, Julien Pettré, and Stéphane Donikian. 2007. Pedestrian reactive navigation for crowd simulation: a predictive approach. In *Computer Graphics Forum*, Vol. 26. Wiley Online Library, 665–674.
- Christopher Peters and Carol O'Sullivan. 2002. Synthetic vision and memory for autonomous virtual humans. In *Computer Graphics Forum*, Vol. 21. Wiley Online Library, 743–752.
- Sara Pizzamiglio, Usman Naeem, Hassan Abdalla, and Duncan L Turner. 2017. Neural correlates of single-and dual-task walking in the real world. *Frontiers in human neuroscience* 11 (2017), 460.
- Prudence Plummer, Sarah Apple, Colleen Dowd, and Eliza Keith. 2015. Texting and walking: Effect of environmental setting and task prioritization on dual-task interference in healthy young adults. *Gait & posture* 41, 1 (2015), 46–51.
- Paphawee Prupetkaew, Vipul Lugade, Teerawat Kamnardsiri, and Patima Silsupadol. 2019. Cognitive and visual demands, but not gross motor demand, of concurrent smartphone use affect laboratory and free-living gait among young and older adults. *Gait & posture* 68 (2019), 30–36.
- Craig W. Reynolds. 1987. Flocks, Herds and Schools: A Distributed Behavioral Model. *SIGGRAPH Comput. Graph.* 21, 4 (Aug. 1987), 25–34.
- Siobhan M Schabrun, Wolbert van den Hoorn, Alison Moorcroft, Cameron Greenland, and Paul W Hodges. 2014. Texting and walking: strategies for postural control and implications for safety. *PLoS one* 9, 1 (2014), e84312.
- W. Shao and D. Terzopoulos. 2005. Autonomous Pedestrians. In *Proceedings of the 2005 ACM SIGGRAPH/EG Symposium on Computer Animation* (Los Angeles, California). ACM, New York, NY, USA, 19–28. <https://doi.org/10.1145/1073368.1073371>
- Shawn Singh, Mubbasir Kapadia, Glenn Reinman, and Petros Faloutsos. 2011. Footstep navigation for dynamic crowds. *Computer Animation and Virtual Worlds* 22, 2-3 (2011), 151–158.
- Jamie Snape, Jur Van Den Berg, Stephen J Guy, and Dinesh Manocha. 2011. The hybrid reciprocal velocity obstacle. *IEEE Transactions on Robotics* 27, 4 (2011), 696–706.
- David L Strayer, Jason M Watson, and Frank A Drews. 2011. Cognitive distraction while multitasking in the automobile. (2011).
- Daniel Thalmann and Soraia Raupp Musse. 2012. *Crowd Simulation*. Springer Science & Business Media.
- Yu Tian, Yifei Huang, Jibo He, and Kunlin Wei. 2018. What affects gait performance during walking while texting? A comparison of motor, visual and cognitive factors. *Ergonomics* 61, 11 (2018), 1507–1518. PMID: 30009681.
- John K Tsotsos. 2001. Motion understanding: Task-directed attention and representations that link perception with action. *International Journal of Computer Vision* 45, 3 (2001), 265–280.
- Jur Van Den Berg, Stephen J Guy, Ming Lin, and Dinesh Manocha. 2011. Reciprocal n-body collision avoidance. In *Robotics research*. Springer, 3–19.
- David Wolinski, Ming C Lin, and Julien Pettré. 2016. WarpDriver: context-aware probabilistic motion prediction for crowd simulation. *ACM Transactions on Graphics (TOG)* 35, 6 (2016), 1–11.
- Kyung-Hoon Yu, Jae-Hun Shim, Sung-Dae Choung, and Hye-Seon Jeon. 2015. Effect of using a Cell Phone on Gait Parameters in Healthy Young Adults: Texting and Texting while Listening to Music. *Journal of Korean Society of Physical Medicine* 10, 4 (2015), 25–31.