Foveated Animations for Efficient Crowd Simulation

FLORIN-VLADIMIR STANCU, University of Leeds, United Kingdom
TOMER WEISS, New Jersey Institute of Technology, United States of America
RAFAEL KUFFNER DOS ANJOS, University of Leeds, United Kingdom

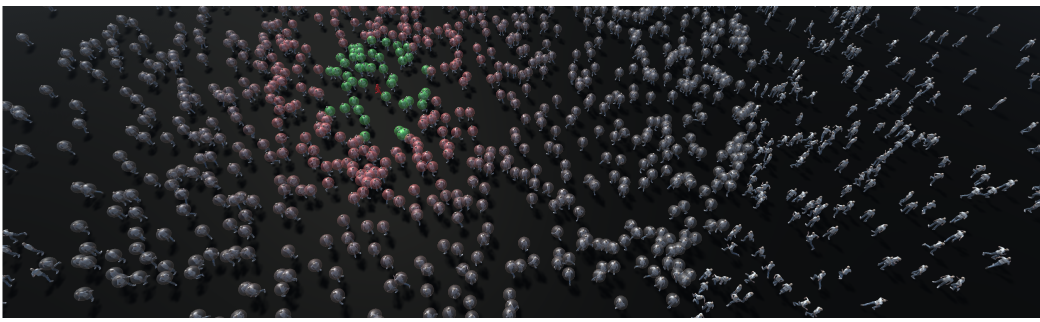


Fig. 1. Crowd simulation where the focus point is depicted in red. Agents in green are inside the foveal region, and fully animated. Red and gray agents have a reduced animation rate according to their distance. Non-highlighted agents are not animated.

Foveated rendering techniques have seen recent development with the advent of commercial head-mounted displays with eye-tracking capabilities. The main drive is to exploit particular features of our peripheral vision that allow optimizing rendering pipelines, which allows using less computational effort where the human visual system may be unaware of differences. Most efforts have been focused on simplifying spatial visual detail on areas not being focused on by adjusting acuity of shading models, sharpness of images, and pixel density. However, other perception pipeline areas are also influential, particularly in certain purpose-specific applications. In this paper, we demonstrate it is possible to reduce animation rates in crowd simulations up to a complete stop for agents in our peripheral vision without users noticing the effect. We implemented a prototype Unity3D application with typical crowd simulation scenarios and carried out user experiments to study subjects' perception to changes in animation rates. We find that in the best case we were able to reduce the number of operations by 99.3% compared to an unfoveated scenario, with opportunities for developments combined with other acceleration techniques. This paper also includes an in-depth discussion about human perception of movement in peripheral vision with novel ideas that will have applications beyond crowd simulation.

CCS Concepts: • **Computing methodologies** → **Animation**; **Perception**; *Virtual reality*; Simulation by animation.

Additional Key Words and Phrases: Perception, Animation, Foveated techniques, Virtual Reality

1 Introduction

The main drivers for development in computer graphics techniques for high performance rendering have always been the video games and entertainment industries. When coupled with the

Authors' Contact Information: Florin-Vladimir Stancu, florinstancu@protonmail.com, University of Leeds, Leeds, West Yorkshire, United Kingdom; Tomer Weiss, tomer.weiss@njit.edu, New Jersey Institute of Technology, Jersey City, New Jersey, United States of America; Rafael Kuffner dos Anjos, r.kuffnerdosanjos@leeds.ac.uk, University of Leeds, Leeds, West Yorkshire, United Kingdom.

popularization of virtual reality (VR) applications using head-mounted displays (HMD), the requirements for interactivity have recently been made even tighter. The extra rendering costs in these devices, increased pixel density, and lower tolerance to latency have pushed researchers into developing specific techniques tailored to these devices. Our understanding of the human visual system (HVS)[Hubel 1982; Tanaka 1996], its statistical behavior in the periphery [Rosenholtz et al. 2012], and its lack of ability to resolve fine detail outside the fovea [Rosenholtz 2016; Strasburger et al. 2011] allowed the development of foveated rendering techniques [Patney et al. 2016; Walton et al. 2021; Wang et al. 2020]. The most popular approaches focus on computing graphics in the periphery using a smaller number of samples [Guenter et al. 2012; Li et al. 2021; Patney et al. 2016] resulting in lower bandwidth and computation in these areas when shading [He et al. 2014] or processing geometry [Murphy and Duchowski 2001]. While this is the most suitable point of optimization for general-purpose applications, some of our understanding of the HVS can be applied to other areas of the pipeline which are stressed in certain contexts.

In this paper we focus on exploring the perception of adaptive animation rates on crowd simulations. We propose the concept of foveated animations; adaptive rate of on-screen animations in the viewer's periphery to achieve the same effect: save system resources without the user experience being impacted. We found that the way our peripheral vision reacts to visual crowding [Rosenholtz 2016; Strasburger 2020] where stimuli are grouped and perceived statistically can be heavily exploited when rendering animated crowds. When thousands of characters need bone transformations, blending, and other factors calculated per frame, one can skip these operations without any impact to the user experience if using the appropriate level of foveation. While work has been done to optimize animation costs [Beacco et al. 2016; Dong and Peng 2019; Toledo et al. 2014], there is yet to be any research on utilizing gaze information to optimize this process without perceived loss of quality.

To explore this approach we implemented a Unity3D prototype of a crowd simulation using typical scenarios, where we adaptively change the animation update rate based on the distance to the focus point. We carried out user studies in both a VR scenario and in a flat screen with a protagonist character to be followed. Using a staircase protocol we determined the validity of our hypothesis, the appropriate levels of foveation for each scenario, and how it relates to our current understanding of the HVS. Our performance analysis shows that we are able to reduce the number of operations to up to 99.3%, which can provide significant efficiency gains in rendering systems for crowd simulation. We also studied the effect of crowding and organized movement in perception of small scale movements in the periphery. We present an in depth discussion and evidence about the role of visual pooling in this context, that is novel in the community to the best of our knowledge.

In summary, our contributions are: (1) An open source Unity3D prototype demonstrating foveated animation in a crowd simulation environment. (2) User studies validating the hypothesis, describing the estimated foveation levels for each scene and setting. (3) Performance evaluation measuring the potential gains of using this approach. (4) Novel discussion on peripheral perception of small scale movements in different settings.

2 Related work

The aim of our work is to exploit knowledge from human perception to accelerate crowd rendering in computer graphics pipelines. This section reviews the related work to these topics in order to proper motivate our hypothesis.

2.1 Perception in Computer Graphics

It is commonly known that, for humans without particular visual impairments, visual acuity is higher in the areas where we focus our eyes. In the central area of our retina lies a region called the *fovea* which is only approximately 5° wide [Strasburger et al. 2011]. Here, we have much higher concentration of photoreceptor cells (cones) responsible for our photopic vision (used in well lit circumstances and allowing color distinction). This falls off significantly outside of the foveal region, while not approaching zero [Anstis 1974]. Similarly, our contrast sensitivity also reduces, due to the lower density of rods which are responsible for our scotopic vision [Legge and Kersten 1987]. This has been used to support the usual description that our peripheral vision is “blurry”, as there is a lower concentration of both cells, thus lower sampling density, and more “interpolation”. Aubert and Föster [Aubert and Förster 1857] stated that our ability to discriminate oriented details scales roughly linearly with eccentricity. This is the notion that was traditionally exploited in popular foveated rendering approaches.

Gaze-contingent graphics have been considered for optimizing rendering pipelines since eye-trackers have been able to provide interactive results. Early on, Levoy and Whitaker [Levoy and Whitaker 1990] explored them in the context of volume rendering, where different mipmap levels could be used to compute the algorithm, based on visual attention. This idea tracks naturally to other rendering pipelines such as raytracing, where the sampling rate of rays can be adjusted according to eccentricity [Fujita and Harada 2014; Wang et al. 2020; Weier et al. 2016]. Simplifying geometry in a gaze-contingent manner was also explored [Murphy and Duchowski 2001; Ohshima et al. 1996], but these works are focused on increasing efficiency using traditional quality trade-offs, not focusing on a close perceptual match.

Guenter et. al [Guenter et al. 2012] introduced the idea that more closely tracks with current popular approaches of foveated graphics in rasterisation pipelines, where sample/pixel density is lower in the periphery. This has been further explored to be content-based with adaptive sampling [Stengel et al. 2016; Vaidyanathan et al. 2014], to counteract our loss in contrast sensitivity [Patney et al. 2016; Tursun et al. 2019], to consider image degradation when seen through HMDs [Hoffman et al. 2018], to cater for eye-dominance. [Meng et al. 2020], and to minimize distortion artifacts in deferred pipelines [Li et al. 2021; Meng et al. 2018].

However on a closer inspection, while the correlation raised by Aubert and Föster [Aubert and Förster 1857] holds and we see with less accuracy, our vision is not simply described as “blurry” by vision science. In fact, one way to describe it is that the visual cortex will use differently sized pooling regions according to eccentricity to process differently oriented stimuli [Carandini et al. 2005; Hubel 1982; Tanaka 1996]. This indicates that the behavior is more statistical in its nature [Rosenholtz et al. 2012], meaning that while we are not able to discern particular details, we can see changes in their statistical distribution [Freeman and Simoncelli 2011; Ziemba and Simoncelli 2021]. An in-depth description of our peripheral vision has been covered in detail in the survey by Strasburger et al. [Strasburger et al. 2011]. These findings motivated future developments in foveated rendering, where statistical behavior in the periphery is reproduced to more closely replicate the frequency content of the original images, [Tariq et al. 2022; Walton et al. 2021] approaching the ventral metamerism studied by Freeman and Simoncelli [Freeman and Simoncelli 2011].

Another way to describe this statistical behavior in the periphery is looking at what is called in vision science the “crowding effect” [Rosenholtz 2016; Strasburger 2020]. In summary, our ability to discern individual elements has not only to do with their size and correlation with the size of the pooling regions in our visual cortex, but highly relates to the surroundings of the element we are trying to distinguish [Greenwood et al. 2009]. This consideration is one the main drivers for the work we present in this paper. It is safe to assume there will be crowding in the periphery in a typical crowd simulation scenario, and omissions in visual detail in the form of movement will be tolerated in the periphery. Supporting this hypothesis, we see the study from Krajancich et al. [Krajancich et al. 2021] describing levels of flicker (i.e. temporal variations in oriented signals) in the periphery that subjects would not perceive. Also Tariq and Didyk [Tariq and Didyk 2024] used oriented Gabor patches in the periphery to correct a mismatch in motion perception introduced by blur. This further supports that we are able to perceive movement in the periphery, but just up to a certain scale, which matches the size of the chosen Gabor patches.

2.2 Virtual Crowds

Our work focuses on speeding-up existing frameworks for visualizing crowds, with the goal of accelerating any existing framework to support real-time, larger scale crowds. For completeness, we first define what virtual crowds are, and then overview related work in this area.

A crowd is a collection of independent, self-actuated agents [Kapadia et al. 2016]. Each agent has individual navigational goals in this shared environment. Agents share the same environment, where they interact and collide with each other. Agent movement is controlled by a navigation algorithm, which needs to ensure that an agent progresses towards its goal, while avoiding collisions [Weiss 2023]. In addition to navigation behaviors, agents may be animated to depict leg, hand, and other body movements to enhance realism of the virtual crowd. Such animations can then be captured and rendered in various styles, ranging from simplified pixels to matching real-world pedestrians. For example, Yin and colleagues used a VR approach to replicate a single person's motion to create a crowd [Yin et al. 2022]. Thus, methods for creating virtual crowds are typically composed of distinct frameworks – simulation [Van Toll and Pettré 2021], animation [Lemonari et al. 2022] and rendering [Beacco et al. 2016], all working together to create the illusion of a realistic, virtual crowd.

For speeding up the visualization of crowds, one popular attempt is to use impostors [Aubel et al. 1998], where complex geometry is replaced by flat billboards, and can be perceived as acceptable in simple scenarios where photorealism and lighting interactions are not crucial. Toledo et al. [Toledo et al. 2014] adapt existing Level-of-detail (LOD) techniques to specific crowd scenarios, using acceleration structures to simplify both geometry and animation chains for elements not in the main plane (e.g. background actors, partially occluded). A more modern approach resorts to instancing of individual elements to minimize the amount of processing [Dong and Peng 2019]. For a detailed review of these techniques we recommend the survey from Beacco et al. [Beacco et al. 2016].

To the best of our knowledge, foveated animation has only been recently explored for fluid simulation [Wang et al. 2024] with success, but not for crowds. We believe crowd simulation differs from this perceptually due to the nature of the stimuli (rigid, oriented motion), but at a larger scale crowd movement is comparable to a fluid. We hypothesise crowds and fluids may have similar visual pooling properties.

3 Methods

We pose the hypothesis that small scale movement (rigged animation updates) can be simplified or omitted entirely if not under the users' foveal vision. We consider here the "critical flicker frequency" (CFF) [Hartmann et al. 1979] where our perception to flicker in the periphery is higher than in the fovea, peaking at 20-50 degrees [Sinha et al. 2017]. When reducing animation rate in the periphery, this may be perceived as flicker by the subjects in immediate peripheral vision. This may be noticeable to users of headsets and displays, and be considered "lower quality". This has recently been revisited in work from Krajancich et al. [Krajancich et al. 2021] proposing an accurate model for different sized spatial variations at different eccentricities. Their work proposes that bandwidth savings can be seven times higher than current spatial foveation models.

While it is possible to design a solution that calculates the spatial frequency of particular animation updates and reduce them accurately to not go over the CFF threshold, we postulate that the crowding effect in our focused scenario (crowd simulations) allows us to take a simplistic approach of reducing animation rate with eccentricity, or not performing animation updates in users peripheral vision, where lack of small scale movement is masked by large scale peripheral movement of crowds, allowing larger reductions of computation time.

This section will outline our design, implementation, and prototype.

3.1 Solution design

Given the target frame-rate R_t of our application we calculate each individual agent target animation rate R_i based on their screen-space distance to the target. The foveal area a_f is discounted from the calculations so all agents in this region are fully animated. Given an agent's position p (center of its bounding box), and the focus point's p_0, and a foveation factor α the *Dynamic foveation* animation rate is calculated as:

$$R_i = \frac{R_t}{max(1, \alpha \cdot ||p - p_o||^2 - a_f)} \tag{1}$$

Previous work [Krajancich et al. 2021] fits a quadratic function in the perception of flicker after passing the critical area. Following their work closely, one would accept introducing flicker near the fovea, but given our context, we cannot decrease animation rates in the fovea, as subjects would be easily aware of the reduction. Our solution matches with their description in that we do not apply any animation reduction inside the foveal area a_f, and model the reduction quadratically past that. Controlling these two variables α and f_a we are able to adjust these parameters depending on the evaluated scene, as the screen-space size of the flickering effect (small scale skeletal animation) will be varied, and one's sensitivity to it is dependant on this [Krajancich et al. 2021]. An example can be seen in Figure 1,

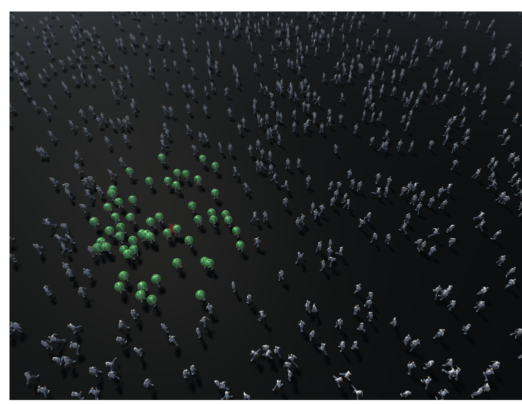


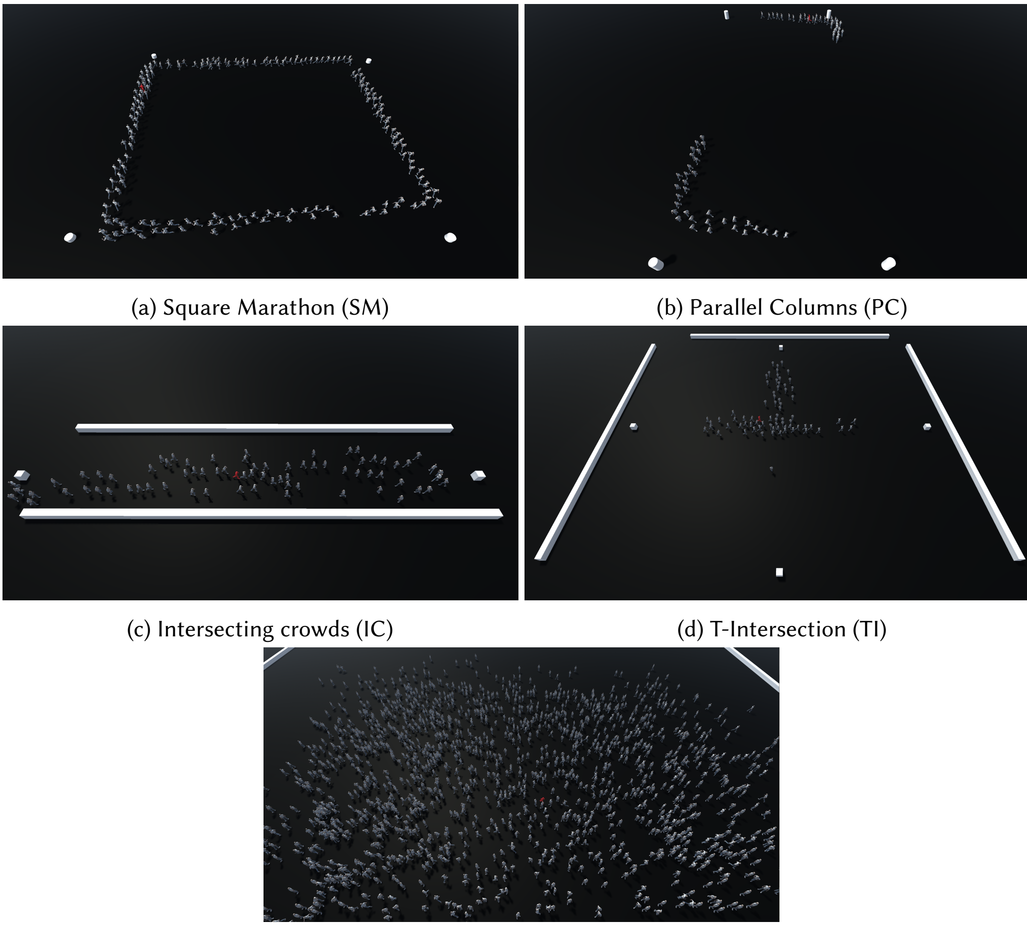
Fig. 2. *Full-stop* behavior, where agents outside the fovea are not animated.

To fully test our hypothesis of completely skipping animation updates for peripheral agents which we will refer to as *Full-stop* for the remainder of this paper, we simply set R_i to be zero outside of a_f, and control a_f as our only foveation factor. An example can be seen in figure 2

3.2 Prototype

To evaluate our hypothesis, we developed a prototype using Unity3D (version 2022.3.19f1, built-in RP), given its ease-of-use for replicability purposes by the academic community, as our main goal is to test our foveated animation hypothesis. The techniques used here can be easily ported to other systems or engines, particularly ones with more advanced crowd simulation mechanisms.

The foveation rate is calculated centrally and set to each agent at the beginning of a frame using equation 3.1. Given that Unity3D takes control of most of the animation updates, we have to opt for manual control with coroutines using the PlayableGraph "Evaluate" at the designated times. On a more flexible rendering system, further optimizations such as batching in evaluation can be easily incorporated.



(a) Square Marathon (SM)

(b) Parallel Columns (PC)

(c) Intersecting crowds (IC)

(d) T-Intersection (TI)

(e) Wander crowd (WC)

Fig. 3. Scenes used by our experiments.

Agent navigation was implemented using Unity's built-in NavMesh system, for a set of different scenes pictured in Figure 3. For the "Wander Crowd" scene, each agent is given a random target in the area and a new one when that is reached, after a brief pause. For all the other scenes, each group of agents is given alternating targets in the scene (represented by cubes or cylinders), creating the desired behaviors. These alternative scenes were included in order to study the effect of our approach with organized movement which is also common in crowd simulation.

Our prototype was extended to VR by using the Unity3D OpenXR SDK and SRAnipal for eye-tracking.

4 Evaluation

We carried out two user studies to evaluate our hypothesis: a structured flat screen test where subjects follow a protagonist character using a chin rest, and an HMD test using an eye tracker. We recruited 12 participants from student and staff body at our institution, all with perfect or corrected vision. We did not need to control for demographics as we did not expect any effect on low-level vision [Shaqiri et al. 2018], and experience with virtual reality applications was not required. Tests were carried out all in the same workstation (Intel(R) Core(TM) i7-14700, 32GB RAM, NVIDIA GeForce RTX 4080 SUPER) and headset to ensure the same viewing conditions for all participants. Ethical approval was obtained through our institution's ethics body (EPS FREC - 2024 2118-2512). All participants were briefed and signed an informed consent form, and were free to stop the experiment at any stage before completion.

4.1 Study 1: Flat screen

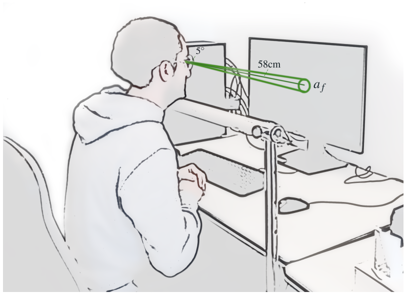
The goal of this first study was to assess in a structured manner the appropriate foveation parameters for each type of scene. We test two different hypotheses here. First, scenes with a higher "crowding effect" with the main example being "Wander crowd" will show higher tolerance for foveation, since it will be less likely that an agents get separated from the group. Secondly, scenes with structured movement will have lower thresholds due to a decreased crowding effect in the movement direction. In a similar way to what was observed in [Ziemba and Simoncelli 2021] where a texture of a different family was more easily detected in our periphery, we aim to see if character animation or the lack of it is more easily spotted in the periphery if the character is part of a unit moving in a similar direction. This study also evaluates the case where a crowd simulation is not being visualized in a VR setting, but has a protagonist agent which is the point of attention for the user.

The study setup can be seen in Figure 4a. Participants used a chin rest placed at 58cm from the screen to ensure similar viewing conditions throughout subjects. First, participants were briefed with a sequence of short videos to demonstrate what a lowered animation rate looks like visually, followed by a foveated and an unfoveated video, where they were free to look around and identify what was the effect they were trying to detect. Participants were instructed to follow the protagonist character (rendered in red), and use the space bar to notify detection of foveation with their peripheral vision, not while looking around.

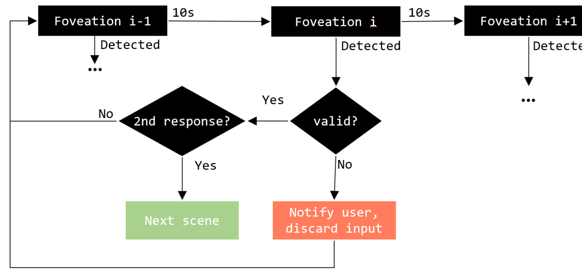
In order to test both our hypotheses, users were exposed to all five scenes twice; once with the *Full-stop* approach and once with the *Dynamic foveation*, with the order alternating with each participant to prevent learning effects. We followed a staircase protocol for this study in order to minimize false positives. This is depicted in Figure 4b. The settings for each test progressed as seen in Table 1, with the foveal area f_a calculated as pictured in Figure 4a spanning 5° of our visual field [Strasburger et al. 2011]. The radius of a_f can be estimated by $r = tan(2.5°) * 58cm \approx 2.5cm$, which then is used relatively to the physical display size to calculate the $a_f \approx 0.072$ which is rounded up to $a_f = 0.08$ (measured as a percentage of the screen) value used for *Dynamic foveation*, as this area will be larger towards the ends of the display.

Each scene started with a control setting of no foveation. Participants signaled seeing the foveation effect by pressing the space bar, and clicking the area where the effect was seen. This was used to detect and discard false positives. When valid detection happened twice at the same setting, subjects were moved to the next scene.

To ensure equal performance and simulation conditions for all participants, all test scenes were simulated and recorded to be displayed to participants during the experiment. Three alternate recordings at each detection were used so "redetecting" the foveation effect did not equate to simply seeing the same stimuli, but a perception of the foveation level itself. All participants were observed by the experiment manager to ensure instructions were being followed and ensure validity of responses. After the experiment participants were informally asked for comments and their impressions on both approaches, and what stood out the most.



(a) Setup for flat screen test. Users used a chin-rest to provide a fixed distance to the display.

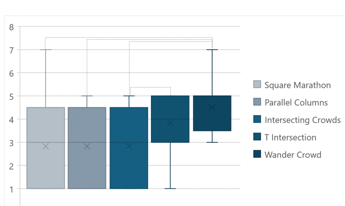


(b) Procedure for staircase protocol depicted for a foveation level $i \in \{0, 11\}$. Each scene was played until foveation was detected twice at the same level, with users notified of false inputs.

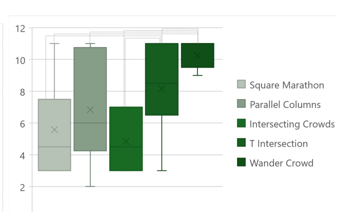
Fig. 4. Flat screen test details

Table 1. Staircase protocol settings. α values used for *Dynamic foveation* with the estimated $a_f = 0.08$, and varying a_f values for the *Full-Stop* approach.

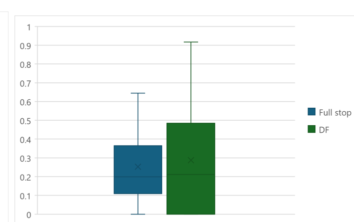
Stage	0	1	2	3	4	5	6	7	8	9	10	11
Dynamic foveation α	0	0.1	0.3	0.5	0.7	0.9	1.1	1.3	1.5	1.7	1.9	2.1
Full-Stop a_f	Inf	0.3	0.25	0.2	0.15	0.1	0.05	0	-	-	-	-



(a) *Full stop results*



(b) *Dynamic foveation*



(c) Distance of foveation detected

Fig. 5. Results of the Flat screen Study. Foveation stages described in Table 1. Connected with a line are pairs with a statistically significant difference. Distance of detection for both approaches in all scenes listed in Normalized screen distance. Details in Section 4.1.1

4.1.1 Results. Results of our study can be seen in Figure 5, which reports the foveation stage in which each effect was noticed. We carried out a pairwise Wilcoxon signed rank test to test for statistical significance, as we did not assume a normal distribution on our data.

For the *Full-stop* approach, we found statistically significant differences between the pairs TI-IC ($p = 0.05$), WC-SM ($p = 0.1$), WC-PC ($p = 0.025$) and WC-IC ($p = 0.01$). For *Dynamic Foveation*, we found statistically significant differences between the pairs TI-SM ($p = 0.05$), TI-IC ($p = 0.005$), WC-SM ($p = 0.005$), WC-PC ($p = 0.005$), WC-IC($p = 0.005$) and WC-TI($p = 0.025$).

Figure 5c shows how far from the protagonist has the user clicked to indicate where the foveation effect was noticed. For both approaches, detection tended to happen closer to the fovea, with a higher spread towards the periphery for the Dynamic Foveation approach.

4.2 Study 2: HMD

After the first study was completed, we carried out an unstructured experiment using an eye-tracked HMD (Vive Pro Eye). The goal of this study was to evaluate the feasibility of this type of application under an eye-tracked scenario where practical limitations such as the speed of the eye-tracker compared to saccadic eye movement could potentially expose the users to characters with imperfect animations for short windows of time.

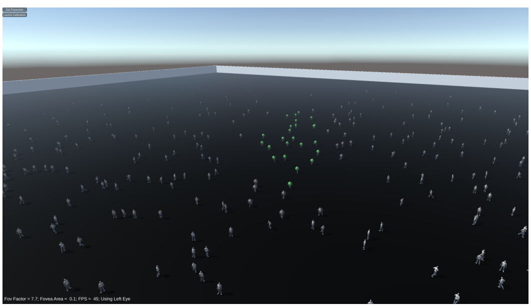


Fig. 6. Average foveation setting during the VR experiment.

After performing eye-tracking calibration, participants were shown the "Wander crowd" scene from a similar point of view to the flat screen test, and allowed to look around freely. Participants were shown the initial foveation settings using the sphere gyzmos seen in Figure 1, also to verify correct functioning of the eye tracker. Dynamic foveation was then increased until participants could tell an effect was happening, and similarly for Full-stop, a_f was decreased until the effect was noticed. At the end of both stages, participants were shown the gyzmos to reveal the effect on the characters in the screen. Values for both steps were registered, and participants were informally interviewed about their experience with the HMD.

4.2.1 Results. Summary of results can be seen in Table 2, depicting the typical stages where subjects said to be comfortable with foveation levels, and no effect was being detected. It's important to note that the scale of values for the VR experiment are distinct from the flat screen test, due to the calculation of $a_f = 0.08$ being dependent on the physical setup used, and distance to the screen.

Table 2. Average values for a_f and α for both approaches.

Full-stop	Dynamic Foveation
0.12	7.7

Figure 6 shows an instance of the VR simulation and its viewpoint with the average settings detected with the user study.

Participants noticed delays in the eye-tracking, which would reveal non-animated characters for brief windows of time. It came up hesitantly most times (e.g. "I thought I had seen something"), as the focus point quickly started animated, and subjects would second guess their perception. When briefed about what was happening, all participants agreed that the delays were not detrimental to a useful user experience of observing crowd behavior.

4.3 Performance Evaluation

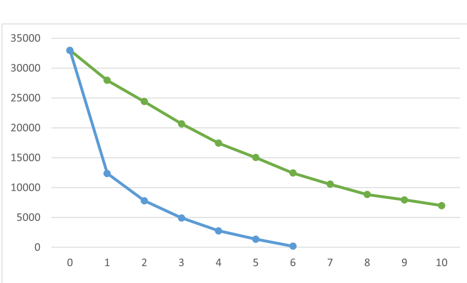


Fig. 7. Animation updates per stage.

We evaluated the performance of the prototype in each of the settings of our staircase protocol seen in Table 1. Results can be seen in Figure 7, for a scene of 1500 charactess, always in view of the camera in our "Wander crowd" scene.

As mentioned in Section 2, many approaches for accelerating both crowd simulations and animation updates exist. We opted for a Unity3D implementation to allow replicability of our study by the community, while being aware of existing limitations on controlling how efficiently animation updates are handled, and uncertainty in the estimated cost of such operations. Thus, we measured number of animation updates, which can be objectively analyzed and applied to more efficient animation systems, and scaled to more complex and detailed rigs.

If we take the median parameter for this scene as a representative value, *Dynamic foveation* reduces the number of animation updates by 78.7%. *Full-stop* will reduce this number by 99.3%. We noticed just a marginal reduction in frames-per-second in our prototype implementation, due to the lack of direct control of object updates.

5 Discussion

The results from both our studies confirm our base hypothesis: foveated animations are a viable approach to reducing costs on rendering of crowd simulations with no impact to user experience. In this section we will discuss in detail the different aspects of our evaluation: Foveation and Crowding, Perception in organized movement, Critical Flicker Frequency, Applications and future work.

5.1 Foveation and Crowding

We found statistically significant results that the tolerated level of foveation in our "Wander crowd" scene is larger than in scenes with organized movement, the exception being WC-TI in the *Full-stop* approach. This confirms both our first hypothesis raised in Section 4.1, that crowding behavior allows us to greatly mask the lack of small scale animations by individual characters. This is a similar effect to what was seen in the perception of textures in a subject periphery under crowding conditions [Ziemba and Simoncelli 2021]. This is novel evidence that we present that this effect is also applicable to the perception of movement. The lack of significance between WC-TI can be explained by groups in TI being more widely spread, making detection of foveation more challenging, thus closer to the thresholds found in WC.

For *Dynamic Foveation*, we are able to take it to the highest setting in our test, which is significantly higher than for all other scenes. Moreover, in the VR study where users were subject to an eye tracker, Figure 6 shows that agents outside of the foveal region are animated at between 5-10Hz (see Figure 1 for a comparison of a normal falloff for foveation), and that was still tolerable by users.

In the *Full-stop* approach, the median foveation level represents $a_f = 0.1$, which is close to our estimate of a_f. Having also in mind the median detection distance of 0.2, we can state that the foveation effect was noticeable mostly on stimuli close to the foveal region, where pooling regions are effectively smaller. Figure 8 shows three cases of foveated agents separated from a group, which led to their detection at between 0.1 and 0.2 nsd. This allows us to recommend that future works should have crowd density as a key variable to determine the foveation factor.

5.2 Perception in organized movement

For our scenes which displayed organized movement, we hypothesized that lower thresholds would be found due to the a decreased crowding effect in the movement direction; i.e. small scale movements and the lack of them would be more noticeable in the periphery, as they are not being masked by the crowd's erratic behavior. We found lower thresholds in these scenes, with TI performed significantly better than IC for the *Full-stop* condition, and than SM and IC for *Dynamic foveation*. While this data may support our hypothesis, our observations from the study trials motivate a distinct reasoning.

Fig. 8. Three cases where detected foveated agents are close to the fovea due to smaller pooling regions.

All subjects were shown a control scene of no foveation as the first step of each trial, and multiple cases of false positives were observed. Following the protocol in Figure 4b, users were notified of this, and restarted the task. This was noted in 9 out of 12 participants for the organized movement tasks, with some repeatedly going through false detections, and expressing out loud their certainty of having seen lack of movement. This indicates that organized movement may induce pooling of its understanding in periphery as a single "large scale" flow. We can draw a parallel to how high frequency content is masked by our lower spatial acuity in peripheral vision, and is popularly explored in foveated rendering approaches [Walton et al. 2021]. Considering the work from Krajanchich et al. [Krajancich et al. 2021], we may state that the smooth animated movement from limbs would fall below CFF. In simpler terms, we may not be able to see this movement or the lack of it in our periphery.

We can thus hypothesise that lower thresholds may have been found as users were not seeing movement in the periphery not because they noticed something missing, but because they are not able to perceive them, and they were instructed to notice that for this experiment. Future experiments should look further into scenes of organized movement using eye tracking software to verify users perception in a similar way to our second study did for a crowd scenario.

5.3 Critical flicker frequency

As seen in Figure 5c detection of *Dynamic foveation* happened still consistently throughout the mid fovea (Q3 = 0.5nsd). Seeing that foveation settings for this approach were still found to be reasonably high, with WC having the median user detecting at the highest setting, we can state that agents being sparsely animated would go beyond CFF which peaks in 20-50 degrees [Sinha et al. 2017].

This was confirmed in our post experiment interviews, where subjects stated flicker was more "disturbing" to the experience, and easily noticed. In a similar fashion, going from being non animated to starting animation in the *Full-Stop* also raised the attention from subjects in the post interview questionnaire, albeit harder to detect as it was a single instance of flicker, after which the character would be smoothly animated.

Looking at the performance gains in using the *Full-Stop* approach, and the challenges that allowing high dynamic foveation to not exceed CFF in our peryphery, we believe it is the best approach for crowd simulation due to its simplicity, performance gains, and having negligible impact on the user experience.

Future work in dynamic foveated approaches should include careful analysis of what is the screen space frequency impact of updating animations for a given object. In scenarios beyond crowd simulation, or when presented with closer perspectives of characters, humans may still have relatively accurate understanding of non verbal communication cues (e.g. faces, hands) due to their anthropological importance. In such scenarios, not animating might be still noticeable in the periphery, and dynamic or selective approaches need to be investigated.

5.4 Applications for crowd simulation

The approach presented in this paper has not only the advantage of decreasing the number of animation update operations, which naturally scale with the number of characters, and the complexity of the 3D model at hand. More interestingly, it allows existing acceleration techniques for crowd simulation to be deployed more efficiently. A smaller number of impostors [Aubel et al. 1998] would be required, considering they can be non-animated in the periphery, and would need a lower update frequency. On a similar note, instancing [Dong and Peng 2019] can be done more agressively by reusing characters with a similar position in their current animation rig. Further studies on foveated LOD could be performed [Toledo et al. 2014], where similar approaches to distance-based LOD could be applied to eccentricity-based techniques.

Beyond pairing foveated animation with the aforementioned methods, future work should explore other scenarios with more complex movement and agent interactions which may be noticeable through our peripheral vision, as organized movement can happen in key scenarios for crowd simulation such as evacuation studies. Better understanding of what type of large scale behaviors will allow masking of animation updates will be key to deploying this approach to a wider array of scenarios.

6 Conclusion

This paper proposes and validates the concept of foveated animations for crowd-simulation scenarios. Through two user studies covering varied scenarios, we demonstrate that it is preferable to not animate characters outside the foveal region, with special attention to parafoveal regions where pooling regions in our visual cortex are still small. Our hypothesis was validated through an experiment with an eye-tracked HMD, where users were not deterred by current limitations of current eye tracking technology. Thus, we demonstrate that it is possible to reduce the number of animation updates in up to 99.3% with no impact to the user experience. We believe this work paves the way for new innovations on acceleration techniques in complex crowd simulation.

This work also provides compelling evidence for further research in perceptual graphics, and our understanding of the Human Visual System. Building on findings of previous works [Krajancich et al. 2021; Sinha et al. 2017], we demonstrate that although counterintuitive, it is better to not animate characters in the periphery due to our increased sensitivity to flicker of a certain frequency in our immediate peripheral vision. Moreover, we verified pooling behavior in the perception of organized movement in our periphery, which has the ability to mask our perception of small scale movement, inducing ventral metamerism between foveated and unfoveated scenes. This novel understanding of the HVS will motivate further research in the field of crowd simulations and a new generation of motion-foveated applications.

Acknowledgments

We would like to thank Jingnan Xu for support with pilot studies, and David Walton for proofreading and discussions at multiple stages.

References

Stuart M Anstis. 1974. A chart demonstrating variations in acuity with retinal position. *Vis Res* 14, 7 (1974), 589–592.

Amaury Aubel, Ronan Boulic, and Daniel Thalmann. 1998. Animated impostors for real-time display of numerous virtual humans. In *International Conference on Virtual Worlds*. Springer, 14–28.

H Aubert and R Förster. 1857. Beiträge zur Kenntniss des indirecten Sehens.(I). Untersuchungen über den Raumsinn der Retina. *Archiv für Ophthalmologie* 3, 2 (1857), 1–37.

Alejandro Beacco, Nuria Pelechano, and Carlos Andújar. 2016. A survey of real-time crowd rendering. In *Computer Graphics Forum*, Vol. 35. Wiley Online Library, 32–50.

Matteo Carandini, Jonathan B. Demb, Valerio Mante, David J. Tolhurst, Yang Dan, Bruno A. Olshausen, Jack L. Gallant, and Nicole C. Rust. 2005. Do we know what the early visual system does? *J Neuroscience* 25, 46 (2005), 10577–10597.

Yangzi Dong and Chao Peng. 2019. Real-Time Large Crowd Rendering with Efficient Character and Instance Management on GPU. *International Journal of Computer Games Technology* 2019, 1 (2019), 1792304.

Jeremy Freeman and Eero P Simoncelli. 2011. Metamers of the ventral stream. *Nature Neuroscience* 14, 9 (2011), 1195–1201.

Masahiro Fujita and Takahiro Harada. 2014. Foveated real-time ray tracing for virtual reality headset. *SIGGRAPH Asia Posters* 14 (2014).

John A. Greenwood, Peter J. Bex, and Steven C. Dakin. 2009. Positional averaging explains crowding with letter-like stimuli. *Proc NAS US* 106, 31 (2009), 13130–13135.

Brian Guenter, Mark Finch, Steven Drucker, Desney Tan, and John Snyder. 2012. Foveated 3D graphics. *ACM Trans Graph* 31, 6 (2012), 1–10.

E Hartmann, B Lachenmayr, and H Brettel. 1979. The peripheral critical flicker frequency. *Vision Research* 19, 9 (1979), 1019–1023.

Yong He, Yan Gu, and Kayvon Fatahalian. 2014. Extending the graphics pipeline with adaptive, multi-rate shading. *ACM Trans Graph* 33, 4 (2014).

David Hoffman, Zoe Meraz, and Eric Turner. 2018. Limits of peripheral acuity and implications for VR system design. *J SID* 26, 8 (2018), 483–495.

David H. Hubel. 1982. Exploration of the primary visual cortex, 1955–78. *Nature* 299, 5883 (oct 1982), 515–524.

Mubbasir Kapadia, Nuria Pelechano, Jan M Allbeck, Norman I Badler, and Norm Badler. 2016. *Virtual crowds: Steps toward behavioral realism.* Springer.

Brooke Krajancich, Petr Kellnhofer, and Gordon Wetzstein. 2021. A perceptual model for eccentricity-dependent spatio-temporal flicker fusion and its applications to foveated graphics. *ACM Transactions on Graphics (TOG)* 40, 4 (2021), 1–11.

Gordon E Legge and Daniel Kersten. 1987. Contrast discrimination in peripheral vision. *J OSA A* 4, 8 (1987), 1594–1598.

Marilena Lemonari, Rafael Blanco, Panayiotis Charalambous, Nuria Pelechano, Marios Avraamides, Julien Pettré, and Yiorgos Chrysanthou. 2022. Authoring virtual crowds: A survey. In *Computer Graphics Forum*, Vol. 41. Wiley Online Library, 677–701.

Marc Levoy and Ross Whitaker. 1990. Gaze-directed volume rendering. In *Proc. i3D*. 217–223.

David Li, Ruofei Du, Adharsh Babu, Camelia D Brumar, and Amitabh Varshney. 2021. A log-rectilinear transformation for foveated 360-degree video streaming. *IEEE Transactions on Visualization and Computer Graphics* 27, 5 (2021), 2638–2647.

Xiaoxu Meng, Ruofei Du, and Amitabh Varshney. 2020. Eye-dominance-guided Foveated Rendering. *IEEE Transactions on Visualization and Computer Graphics* 26, 5 (2020), 1972–1980. doi:10.1109/TVCG.2020.2973442

Xiaoxu Meng, Ruofei Du, Matthias Zwicker, and Amitabh Varshney. 2018. Kernel foveated rendering. *Proc. i3D* 1, 1 (2018), 1–20.

Hunter Murphy and Andrew T Duchowski. 2001. Gaze-contingent level of detail rendering. *Proc. Eurographics* (2001).

Toshikazu Ohshima, Hiroyuki Yamamoto, and Hideyuki Tamura. 1996. Gaze-directed adaptive rendering for interacting with virtual space. *Proceedings - Virtual Reality Annual International Symposium* (1996), 103–110.

Anjul Patney, Marco Salvi, Joohwan Kim, Anton Kaplanyan, Chris Wyman, Nir Benty, David Luebke, and Aaron Lefohn. 2016. Towards foveated rendering for gaze-tracked virtual reality. *ACM Trans Graph* 35, 6 (2016), 179.

R Rosenholtz. 2016. Capabilities and Limitations of Peripheral Vision. *Annual review of vision science* 2 (2016), 437.

Ruth Rosenholtz, Jie Huang, Alvin Raj, Benjamin J. Balas, and Livia Ilie. 2012. A summary statistic representation in peripheral vision explains visual search. *J Vision* 12, 4 (2012), 14–14.

Albulena Shaqiri, Maya Roinishvili, Lukasz Grzeczkowski, Eka Chkonia, Karin Pilz, Christine Mohr, Andreas Brand, Marina Kunchulia, and Michael H. Herzog. 2018. Sex-related differences in vision are heterogeneous. *Scientific Rep* 8, 1 (2018), 7521.

Raunak Sinha, Mrinalini Hoon, Jacob Baudin, Haruhisa Okawa, Rachel OL Wong, and Fred Rieke. 2017. Cellular and circuit mechanisms shaping the perceptual properties of the primate fovea. *Cell* 168, 3 (2017), 413–426.

Michael Stengel, Steve Grogorick, Martin Eisemann, and Marcus Magnor. 2016. Adaptive image-space sampling for gaze-contingent real-time rendering. 35, 4 (2016), 129–39.

Hans Strasburger. 2020. Seven Myths on Crowding and Peripheral Vision. *i-Perception* 11, 3 (2020).

Hans Strasburger, Ingo Rentschler, and Martin Jüttner. 2011. Peripheral vision and pattern recognition: A review. *J Vision* 11, 5 (2011), 13–13.

Keiji Tanaka. 1996. Inferotemporal Cortex and Object Vision. *Ann Rev Neuro* 19, 1 (1996), 109–39.

Taimoor Tariq and Piotr Didyk. 2024. Towards Motion Metamers for Foveated Rendering. *ACM Transactions on Graphics (TOG)* 43, 4 (2024), 1–10.

Taimoor Tariq, Cara Tursun, and Piotr Didyk. 2022. Noise-based enhancement for foveated rendering. *ACM Transactions on Graphics (TOG)* 41, 4 (2022), 1–14.

Leonel Toledo, Oriam De Gyves, and Isaac Rudomín. 2014. Hierarchical level of detail for varied animated crowds. *The Visual Computer* 30 (2014), 949–961.

Okan Tarhan Tursun, Elena Arabadzhiyska-Koleva, Marek Wernikowski, Radosław Mantiuk, Hans-Peter Seidel, Karol Myszkowski, and Piotr Didyk. 2019. Luminance-contrast-aware foveated rendering. *ACM Trans Graph* 38, 4 (2019), 1–14.

Karthik Vaidyanathan, Marco Salvi, Robert Toth, Tim Foley, Jim Akenine-Möller, Tomas Nilsson, Jacob Munkberg, Jon Hasselgren, Masamichi Sugihara, Petrik Clarberg, Tomasz Janczak, and Aaron Lefohn. 2014. Coarse Pixel Shading. In *Proc. HPG*.

Wouter Van Toll and Julien Pettré. 2021. Algorithms for microscopic crowd simulation: Advancements in the 2010s. In *Computer Graphics Forum*, Vol. 40. Wiley Online Library, 731–754.

David R. Walton, Rafael Kuffner Dos Anjos, Sebastian Friston, David Swapp, Kaan AkÅŸit, Anthony Steed, and Tobias Ritschel. 2021. Beyond Blur: Ventral Metamers for Foveated Rendering. *ACM Trans. Graph. (Proc. SIGGRAPH 2021)* 40, 4 (2021).

Lili Wang, Runze Li, Xuehuai Shi, Ling-Qi Yan, and Zhichao Li. 2020. Foveated Instant Radiosity. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Porto de Galinhas, Brazil, 1–11. doi:10.1109/ISMAR50242.2020.00017

Yue Wang, Yan Zhang, Xuanhui Yang, Hui Wang, Dongxu Liu, and Xubo Yang. 2024. Foveated Fluid Animation in Virtual Reality. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. IEEE, 535–545.

Martin Weier, Thorsten Roth, Ernst Kruijff, André Hinkenjann, Arsène Pérard-Gayot, Philipp Slusallek, and Yongmin Li. 2016. Foveated real-time ray tracing for head-mounted displays. 35, 7 (2016), 289–298.

Tomer Weiss. 2023. Fast position-based multi-agent group dynamics. *Proceedings of the ACM on Computer Graphics and Interactive Techniques* 6, 1 (2023), 1–15.

Tairan Yin, Ludovic Hoyet, Marc Christie, Marie-Paule Cani, and Julien Pettré. 2022. The one-man-crowd: Single user generation of crowd motions using virtual reality. *IEEE Transactions on Visualization and Computer Graphics* 28, 5 (2022), 2245–2255.

Corey M Ziemba and Eero P Simoncelli. 2021. Opposing effects of selectivity and invariance in peripheral vision. *Nature communications* 12, 1 (2021), 4597.