



DEGREE PROJECT IN TECHNOLOGY,
FIRST CYCLE, 15 CREDITS
STOCKHOLM, SWEDEN 2019

Impact of Crowd Density and Camera Perspective on User Sensitivity Towards Impostor Resolution Degradation in Crowd Simulators

KTH Bachelor Thesis Report

Nils Lindberg Odhner and Carl Johan
Freime

Inverkan av densiteten av folkmassan och kameraperspektiv på användarkänslighet mot försämring av impostor- upplösning i folkmassimulering

Nils Lindberg Odhner

Carl Johan Freme

Degree Project in Computer Science, DD142X

Date: June 2019

Supervisor: Christopher Peters

Examiner: Örjan Ekeberg

KTH Royal Institute of Technology

School of Electrical Engineering and Computer Science

Abstract

The entertainment industry and research departments feature many applications of crowd simulation. Typically, it does not suffice to simply generate crowds large in number; they must also be realistic in terms of behaviour and visual fidelity, which is far from an easy or computationally cheap task. Because of this, methods to simulate crowds in a realistic manner while keeping complexity at a manageable level must be explored and implemented.

Implementing impostors in crowd simulation is a ubiquitous technique for managing rendering complexity. Impostors, 2D image representations of virtual 3D objects, improve performance due to their relative simplicity, and can be implemented with varying resolution depending on performance needs; the lower the image resolution of impostors, the higher the efficiency. Not all crowds are the same, some are denser, while others are sparse. On the same note, the scenarios, and thus viewing perspectives, in which crowds are found vary as well; some crowds are seen at eye-level, others from a bird's-eye view. This thesis explores how viewers' sensitivity towards impostor resolution degradation is affected by crowd density and perspective. This is done by conducting a user study in which participants are shown a number video clip pairs. Each pair features clips showing the same crowd scenario, but with one clip only containing high-resolution impostors, and one clip with a gradually increasing share of *low-resolution* impostors. For each video clip pair, participants are tasked with identifying the one clip containing low-resolution impostors as quickly as possible. The video clip pairs feature different combinations of crowd density and viewing perspective, and the results from the study show that crowd density has a significant effect on viewers sensitivity towards impostor resolution degradation. Although perspective seems to have an indicative effect, it was not found to be significant.

Keywords

Impostors, crowd simulation, thesis

Sammandrag

Dagens underhållningsindustri och forskningsvärld har många tillämpningar av folkmasssimulering. Vanligtvis räcker det inte att bara simulera folkmassor stora till antalet individer, de måste också vara realistiska vad gäller beteende och visuell trovärdighet, vilket är långt ifrån en lätt eller beräkningsmässigt billig uppgift. Av denna anledning måste metoder för att simulera folkmassor på ett realistiskt sätt, samtidigt som man behåller komplexiteten på en hanterbar nivå, undersökas och implementeras. Denna rapport undersöker och diskuterer i vilken grad detaljnivån hos simulerade folkmassor kan sänkas genom användning av impostors, 2D-representationer av 3D-objekt vilka vanligtvis används när objektet är för långt bort för att betraktare ska kunna särskilja en 3D-modell av objektet från en bild av samma objekt. Alla folkmassor är inte desamma, vissa är tätare, medan andra är glesa. På samma sätt skiljer sig de scenarion, och därmed perspektiven, i vilka massorna återfinns; vissa folkmassor ses på ögonnivå, andra från ett fågelperspektiv. Det undersökta problemet i denna rapport gäller hur tittarnas känslighet mot försämring av impostorupplösning påverkas av folkmassans densitet och perspektivet från vilket de betraktas. Detta görs genom att genomföra en användarstudie där deltagare visas ett antal videoklipp-par. Varje par innehåller klipp som visar samma folkmassescenario, men där endast ett klipp innehåller uteslutande högupplösta impostors, och ett klipp med en gradvis ökande andel av impostors med låg upplösning. För varje videoklipp-par har studiedeltagarna i uppgift att identifiera det klipp som innehåller lågupplösta impostors så snabbt som möjligt. Videoklipp-paren har olika kombinationer av folkmassodensitet och visningsperspektiv, och resultaten från studien visar att folkmassodensitet har en signifikant effekt på tittarnas känslighet mot försämring av impostorupplösning. Även om perspektiv verkar ha en indikativ effekt, har den inte visat sig vara signifikant.

Nyckelord

Kandidatexamensarbete, simulering av folkmassor, impostors

Acknowledgements

We the authors would like to dedicate this section to personally thank Jack Shabo for his work on the crowd simulator used in this study, Jacob von Eckermann for sharing his lessons learned when conducting a similar study, and Christopher Peters for guidance, continuous feedback, suggesting related works and providing essential resources.

Authors

Nils Lindberg Odhner nodhner@kth.se and Carl Johan Freme freme@kth.se
The School of Electrical Engineering and Computer Science
KTH Royal Institute of Technology

Place for Project

Stockholm, Sweden

Examiner

Örjan Ekeberg
KTH Royal Institute of Technology

Supervisor

Christopher Peters
KTH Royal Institute of Technology

Contents

1	Introduction	1
1.1	Problem statement	2
1.2	Approach	2
2	Background	4
2.1	Crowd simulation	4
2.2	Impostors	6
2.3	Scene and screen center bias	7
3	Method	9
3.1	Implementation	9
3.2	Evaluation	12
3.3	Demographic makeup	14
3.4	Hypotheses	14
4	Result	16
4.1	Mean response distance and response precision	16
4.2	Test of hypothesis H_0	17
4.3	Test of hypothesis H_1	18
5	Discussion	19
5.1	Analysis of results	19
5.2	Limitations	20
6	Conclusions	24
7	References	25

1 Introduction

Virtual crowds of different kinds appear in a wide variety of scenarios, either for entertainment purposes within film, TV and video games, or for research purposes to gain a better understanding of how crowds behave in different situations. To mention just a few examples, crowd simulation software enables us to visualize and study animal herd migrations, large pedestrian crowds in cities, as well as the flow of people in a building evacuation scenario. As computing technology advances, simulation of larger and more life-like crowds is made possible, but this is accompanied by a higher threshold for what is generally considered a ‘realistic’ crowd. The computational challenge of generating large, believable crowds efficiently can be countered with implementing *Level of detail (LOD)* techniques. LOD is a means of optimization through lowering the visual fidelity of objects in a scene for the sake of rendering efficiency. A scene may be (and probably is, when possible) populated with objects of different levels of detail, and objects’ LOD in a scene might range from 3D models rendered with (the screen’s) full resolution down to 2D representations of a considerably lower resolution. A *high* LOD is preferably implemented for objects occupying a large portion of the screen view, or if close to the camera, since these objects would be under the viewer’s scrutiny, more so than distant, less significant or small objects. Thus, we can afford to represent some objects with a relatively *low* LOD, without significantly worsening the viewing experience as whole, but possibly reaping performance gains.

A common example of utilizing LOD techniques, is the use of *impostors*, which are 2D representations of 3D objects. Implementing an impostor system within crowd simulators is a proven way of improving performance; using *texture compositing*, impostors allow for quicker rendering and efficient storing of the impostors texture data.[6]

Impostors can have differing levels of image detail (resolution), in between them as well. This begs the question of how far one can go in terms of lowering impostors’ resolution without contributing to a viewing experience of obviously lesser quality.

With this study, then, the goal is to discover how sensitive viewers are to degrading

detail levels within crowds of impostors, given different conditions.

1.1 Problem statement

Different applications of crowd simulation implies different combinations of conditions in terms of several variables, the most obvious being crowd density and size, and camera positioning, that is to say, the perspective from which the crowd is viewed.

Knowing that the use of impostors improves performance, it is worthwhile to explore how these variables together might affect the perceived fidelity (and thus probably believability) if one renders impostors at degrading levels of resolution, for the sake of performance. Can we become more aware, and take advantage of whether these variables allow for lower resolutions? Which of these variables affect the viewers sensitivity to resolution variations the most? The research question at hand is, then

How crowd density and camera positioning affect viewer sensitivity of impostor resolution degradation?

Any possible performance gains made through impostor resolution degradation have to be weighed against perceived loss of visual fidelity for the viewer, whatever the *stimuli*, the scene, may be. Depending on the stimuli settings, in terms of the same variables as the ones considered in the 3D-to-2D case, sensitivity to resolution degradation may vary. A higher density of agents, for example, limits the level of detail possible with regards to sustainable performance, but it also reasonably allows for a higher degree of resolution degradation, due to occlusion of distant agents.

1.2 Approach

For this study, it was decided upon to consider two main variables and how they may affect users' perception of stimuli fidelity.

To explore the implications of user perception when using impostors of varying LOD in simulated crowds, a user-study will be conducted. Study participants will

be exposed to a number of video clip *pairs*, where each pair has the same certain combination of crowd density and camera positioning values.

During each of these clips, a crowd of impostors will be walking in a broad lane towards the viewer. Depending on a threshold camera distance value, some impostors will be rendered in a *high* resolution, while others, whose distance exceeds the threshold, will be rendered in a considerably lower resolution. As a clip progresses, this threshold will increase, thus resulting in an ever-higher proportion of *low*-resolution impostors in the scene, and participants will be prompted to take note by pressing an on-screen button whenever they identify the clip in which there are low-resolution impostors.

2 Background

This section details the background of the main concepts and techniques discussed and implemented throughout the study.

2.1 Crowd simulation

Crowd simulation entails simulating the behaviour and look of large groups of different types of agents in varying settings[2]. The purpose of crowd simulator software, then, is to perform this task of computing and presenting the virtual crowd in an efficient manner. Progress within the field of crowd simulation directly affects the possibilities to create believable experiences of large crowds within for example movies and video games, as well as in science in general, as we try to further understand how entities, such as people, behave in real-life crowd situations.

Simulating realistic crowds is a challenging task; the developer needs to balance performance and the believability of the crowd. The user's perceived believability of a simulated crowd depends on the level of detail (LOD) in behaviour and rendering, both of which affect performance. Consequently, it is desirable to find efficient methods for handling both aspects, a couple of which will be described in the following sections.

2.1.1 Agent density in crowd simulators

Many, if not most applications of crowd simulation implies crowds containing hundreds or thousands of individual agents. The need for visualizing large number of agents on screen is typically accompanied by restrictions in agent movement, in part with respect to the environment (buildings, roads etc.), and in part with respect to the fellow surrounding agents. The number of agents and the specifics of surroundings affect a crowd's density, which may be either relatively sparse or relatively dense. This density variable might affect the visual perception of the crowd. For example, sparser crowds have been found to allow for higher

detection rates of anomalies in the crowd. More dense crowds are harder to scan for agent anomalies; viewers tend to let their gaze scan individual agents rather than the crowd as a whole, and with higher agent density the chances of finding anomalies in individual agents decrease.[4]

2.1.2 Viewing perspective in crowd simulators

As described earlier in this chapter, applications have different needs for how simulated crowds are presented, not least in terms of perspective, i.e. the distance and angle from which the crowd is viewed. Viewing a crowd from street (eye) level is a different experience from seeing the same crowd from a top-down view. Whether a crowd is seen from above or at street level has implications regarding what crowd agents are at all visible at any given moment, and also affects how close individual agents can get to the camera. This also means that perspective puts limitations on how large share of the screen a single agent can occupy, which in turn affects what level of scrutiny we can expect from viewers on agents. Lastly, the more screen space occupied by individual agents, the less background agents will be visible. One 2018 study, found that the viewing distance from which agents are seen has an effect on viewers' ability to identify agents of lower image detail. The study only compared two agents of different quality (3D model v. 2D impostor) walking next to each other, however, and viewing perspective in terms of angle was not found to be significant.[1]

2.1.3 UIC and crowd simulators

As for the behavioural aspect of crowd simulation, different algorithms have been presented through the years. The Unilateral Incompressibility Constraint (UIC) is one method of addressing the challenge of making virtual crowds move in a both believable and efficient manner. It was first introduced by Narain et al. in and works well for high-density crowds with a high number of agents[5]. It utilizes fluid dynamics as a means to achieve a both realistic and computationally efficient way of simulating the flow of agents in a crowd.

2.2 Impostors

When it comes to rendering large groups of objects, efficiency can be, and is often, achieved by using impostors, which are less costly to render, thus computationally less taxing. Impostors (also called stand-ins) are image-based LODs which approximate the geometric detail of 3D models with the image detail of a 2D texture representation. The image is mapped onto a flat polygon, which is generally directed towards the camera.[6]

Typically, objects such as virtual characters are not static; they are animated and are viewed from different directions. This means that, in order to represent such a 3D object in 2D, images of the object must exist for several angles and animation frames. The more angles (possibly around both the horizontal and vertical axis) we consider capturing for each animation frame, the more texture memory is consumed. With this in mind, one usually settles for a fixed number of angles around the axes.

As our crowd simulator runs, the relative angle between camera and object is calculated, with which the corresponding generated impostor image can be chosen for rendering. Many impostor images are typically gathered onto a single texture map for better efficiency (see section 2.3.2 on texture compositing). Determining the correct impostor image based on both viewing angle and the animation frame entails some complexity. To mitigate this, one might opt for a lower (compared to the original 3D model) temporal resolution for the impostors, that is to say, the rate at which the animation frame is updated is lowered.[6]

2.2.1 Impostors and LOD thresholds

Using LOD begs the question of when to switch between lower and high resolution models. The threshold for when to render impostors in place of 3D agents can be determined in several ways, including camera distance, occupied screen space and arbitrary priority levels, with the first-mentioned criterion being the easiest to implement.[6]

Analogous to improving performance through rendering 2D impostors in place of 3D agents, it can be beneficial to render impostors at degrading levels of detail.

Just as the crowd simulation modeller must choose a breaking-point for when to render impostors over 3D agents depending on different environment variables, one might decide upon a point where impostors of lower resolution should be rendered in place of higher resolution ones. For example, if impostors (rather than 3D agents) of a certain resolution are to be rendered starting from a certain distance, impostors of a resolution half as high as the first ones might be rendered starting from twice the first distance instead, which would reasonably improve performance. From here on, this breaking-point distance is what will be referred to as the *LOD threshold*.

In conclusion, then, crowd simulation can imply considerable levels of complexity, first of all in terms of how to simulate a believable crowd with adequate behaviour and rendering techniques implemented through various algorithms. Furthermore, simulating crowds becomes an even more complex matter as one typically must consider performance requirements - calculating proper behaviour and rendering can be demanding work, memory-wise and computationally.

2.2.2 Impostors and texture compositing

Impostors deal with 2D images; this allows for the use of certain techniques for optimizing how texture data is stored and handled. By gathering multiple subtextures on a single texture map (an image file), the number of state changes each frame can be reduced. For example, one texture map might contain the subtextures of all animation frames of a character. Whatever part of the animation is to be rendered, for a given viewing angle of the impostor, we need only look within the single texture map.[6]

2.3 Scene and screen center bias

Since this thesis will be based on a user study involving visual perception of screen stimuli, it is worth mentioning research findings on the center bias of users' gaze; Bindemann states in a 2010 article that there is a tendency for viewers' gaze to focus on a screen's or scene's center [9]. In studies relying on

users' visual perception, this may lead to areas of whatever the stimuli to receive disproportionate amounts of attention and scrutiny from the viewer, leading to misrepresentative results. This has at least a couple of possible implications. Firstly, this bias could potentially be taken advantage of to develop efficient crowd simulators; knowing that viewers' gaze typically wander towards certain parts of the image before them, one could heighten the level of detail of whatever part of the crowd is within that area of the image, while allowing the rest of the crowd to be presented with less detail, since less attention is directed there. Combining this knowledge with eye-tracking technology could be a powerful means of achieving efficiency.

Secondly, because the approach of our user study entails moving a LOD ever closer to the viewer threshold (see Method or Approach section), being aware of the above bias will be useful when examining the study results. If most study participants identify said threshold at the center of the stimuli, this could potentially, in part or fully, be due to a center bias.

3 Method

Finding an answer to our problem statement implied a two-step process. The evaluation step, that is the actual user study, was designed to show participants a set of video clips, which feature an environment of a simulated crowd of impostors. The creation of this user study stimuli took place in the implementation step, and is detailed first.

3.1 Implementation

Having decided upon a research question, a crowd simulator implementation which satisfies our need of impostors were built. This section describes the project's prerequisites and how they were built upon to create the environment and material ultimately needed for the user study.

The systems written for this project was based on an existing crowd simulator, implemented through the Unity 3D game engine, developed by Jack Shabo [4] as part of a MA thesis. The software supports crowds of up to 3000 animated 3D agents and achieves state of the art crowd behaviour and performance through implementation of UIC (see previous chapter), combined with the use of groups. Shabo's simulator acknowledges that people in crowds often move in small groups, a fact which, when implemented in this kind of system, results in heightened realism and better performance. Shabo's work provided a foundation on which only a system for generating and rendering impostors in place of the 3D agents was needed. Algorithms for crowd behaviour were left untouched.

The implementation of the impostor system consists of two main parts; capturing and rendering. Capturing, concerning the creation of impostor images. Rendering, concerning how the images ought to be displayed.

The video clip sequences used in the evaluations were created in Unity 3D, a real-time game engine used for creating and running games and other software, such as simulations. C is the standard programming language used for Unity applications, and the bulk of the implementation is written in C, however one supporting algorithm in the capturing phase of the implementation was made with

Python.

3.1.1 Capturing

As earlier stated, when implementing an impostor system one has to consider both performance and authenticity of the impostors. On the one hand, having fewer impostor textures, achieved by capturing images from fewer angles, will save memory. But on the other hand, this results in less convincing impostors due to the transitions not being perceived as smooth as the relative angle between the impostor and camera changes. In the following implementation, it was concluded that having steps of 10 degrees between each shot was appropriate. The amount of aggregate space was left manageable and it was deemed that the transitions were hard enough to spot. Furthermore, the participants of the study were only shown the impostors facing forward, therefore snapshots were only taken from the front of character models. Hence, if looked at with spherical coordinates, with the model in the center and the camera on the x-axis, the zenith-angle varied between -60 and 60 degrees while the polar angle varied between 0 and 90 degrees. For each combination of angles and animation frame, one high resolution image was created, and one low resolution was generated. The high- and low-resolution image has a resolution of 512 x 1024 and 128 x 256, respectively.

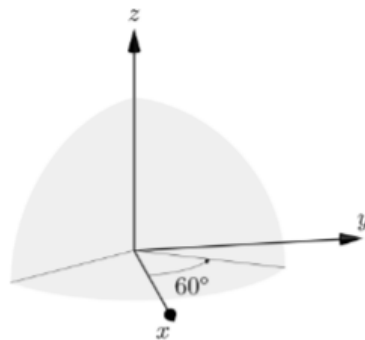


Figure 3.1: *Illustration visualising the angles from which impostors are seen from by participants. The impostor stands along the z-axis, faced along the x-axis.*

A system was developed to iterate over all 108 permutations of angles capturing a snapshot of each unique animation frame. Thereafter, with the help of a Python

script, the animation frames sharing the same angle pair were gathered onto a single 4096x8192 pixel texture map.

3.1.2 Rendering

To begin with, the earlier created textures are used to create Materials, an object in Unity that defines how a surface (2D or 3D) should be rendered with regards to texture, shading/lighting, etc. Each material corresponds to a certain pair of impostor-to-camera angles and a certain animation key frame.

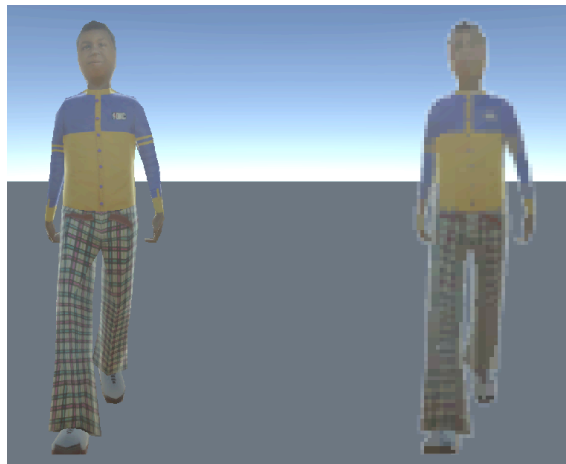


Figure 3.2: *Image of two impostor agents. High resolution on the left, low resolution on the right.*



Figure 3.3: *Image of two impostor agents. High resolution on the left, low resolution on the right.*

The impostor image material is displayed on a quad, a flat 2D Unity plane object. This quad is, in turn, always faced towards the camera, and so the viewer's gaze

is perpendicular to whatever quad, and thus, impostor image, it may be set on. If the relative angle between the camera and impostor changes significantly, the material changes accordingly, which will also inherit wherein the animation the impostor was, so that the animation does not restart abruptly.

3.2 Evaluation

In order to determine to what degree users are able to distinguish the presence of low-resolution impostors in a crowd, a user experiment was designed. More specifically, the purpose of the user study is to examine viewers' sensitivity to the use of impostors as a LOD technique. By exposing viewers to crowd scenarios containing impostors of high and low resolution, in which the distance-based threshold for when to switch between the two detail levels gradually decreases.

Using impostors of two resolution levels, low and high, an attempt to measure said sensitivity within a scene containing a simulated crowd, was done by exposing viewers to a gradually decreasing threshold, which determines when to switch between the two detail levels. In essence, viewer sensitivity was be measured by gradually increasing the number of low-resolution impostors before the eyes of the study participants. Study participants were shown eight video clips, each featuring two simulated crowds moving towards the camera. The possible configurations of crowd density and camera positioning are found in Table 3.1 below. One crowd contained impostors of lesser resolution while the other only contained high resolution impostors. Each clip was 65 seconds long, and as a clip progresses, the LOD threshold is moved closer to the viewer's camera every 5 seconds. With the far end of the crowd being at a distance of D length units from the camera, the LOD threshold is initially placed at 100% of D units from the camera.

As participants viewed a clip, they were asked to press one of two buttons whenever they identified degradation of impostor resolution. One button for the crowd containing low-resolution impostors and one for the crowd containing only high-resolution impostors. The participants were asked to click on the

button that corresponds to the crowd they perceived contained low-resolution impostors.

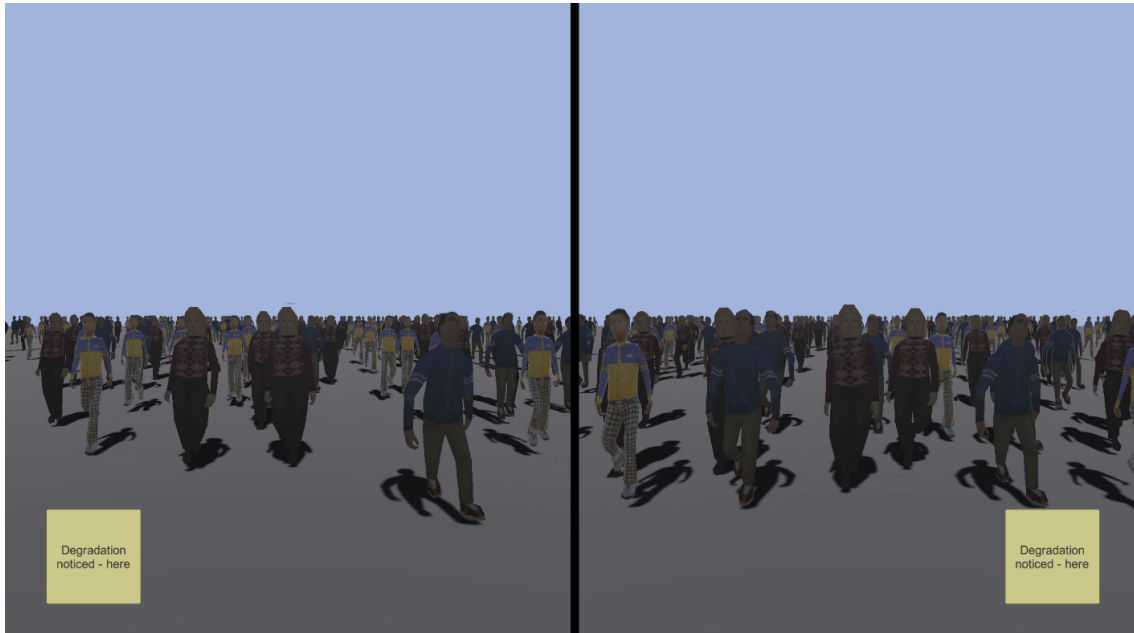


Figure 3.4: *Showing one of the stimuli variations. In this case, high density and street-level view.*

In this way, data on viewers' sensitivity to impostor resolution degradation is gathered, allowing us to answer the problem statement of this thesis. For each participant and each video-clip pair, the result is comprised of one data point for when the user perceive they can spot a difference and one data point indicated if they were correct with their assessment.

The participants viewed the clips in full-screen on a 21 inch monitor with a 1920x1080 screen resolution.

Crowd density	Camera angle
High	Street level (0°)
High	Top-down (45°)
Low	Street level (0°)
Low	Top-down (45°)

Table 3.1: Stimuli variations in the user study

3.3 Demographic makeup

Twelve participants took part in the study. Their age varied from 23 to 28 and 1/3 identified themselves as female and 2/3 as male. All study participants estimated their vision at the time of the experiment to be 4 out of 5 or higher. The information was gathered in a questionnaire before the study, see appendix B.

Furthermore, each study participant saw the same combination of stimuli twice, which in turn resulted in 24 data points for each stimuli.

3.4 Hypotheses

It was decided upon to use a significance level of 5%. For each variation of stimuli a hypothesis testing is formulated in terms of three hypotheses. The null hypothesis says that, with a 95% confidence, users perceive themselves unable to identify crowds containing low-resolution impostors. The first alternative hypothesis says that, with a 95% confidence, users perceive themselves able to identify crowds containing low-resolution impostors. A *two-way ANOVA* analysis was performed to determine if the users' perceived an ability to distinguish low-resolution impostors.

The data gathered a measurement of precision, that is, to what degree participants correctly assess what crowds contain low-resolution impostors. This results in precision rates for all stimuli variations, and so, within a 95% confidence interval, if the difference in precision rates between stimuli variations is separated from 0, one can conclude that users have an actual ability to identify low fidelity impostors. This is the second alternative hypothesis.

Density

H_0 : Density does not have a significant effect on users' perceived ability to identify low fidelity impostors.

H_1 : Density has a significant effect on users' perceived ability to identify low fidelity impostors.

H_2 : Density has a significant effect on users' actual ability to identify low fidelity impostors.

Perspective

H_0 : Perspective does not have a significant effect on users' perceived ability to identify low fidelity impostors.

H_1 : Perspective has a significant effect on users' perceived ability to identify low fidelity impostors.

H_2 : Perspective has a significant effect on users' actual ability to identify low fidelity impostors.

Density and perspective interaction

H_0 : Density and perspective interaction does not have a significant effect on users' perceived ability to identify low fidelity impostors.

H_1 : Density and perspective interaction has a significant effect on users' perceived ability to identify low fidelity impostors.

H_2 : Density and perspective interaction has a significant effect on users' actual ability to identify low fidelity impostors.

4 Result

4.1 Mean response distance and response precision

In figure 4.1 below, one can observe the mean distance required for the participants to notice the resolution degradation of the impostors depending on the combination of stimuli variables. The error bars represent the standard deviation of each category. See table C.1 in appendix C. One can see a marginal difference between high-and low-density stimuli.

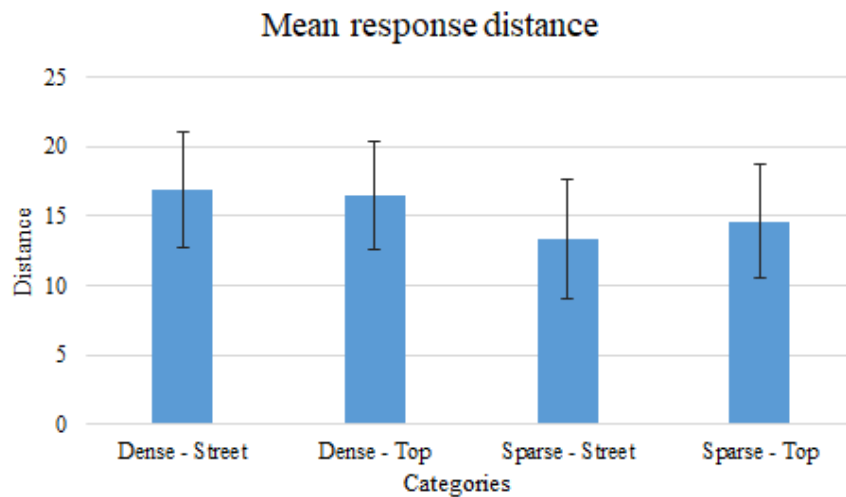


Figure 4.1: *Mean response distance with standard deviation. Distance is in terms of unity units between the camera and the low quality threshold. With permutations of variables - density and perspective - as categories.*

Figure 4.2 present how well the participants were able to asses which stimuli had impostors of low-resolution impostors. See table C.4 in appendix C. There is a small indication that both variables have an impact on the precision of the study participants.

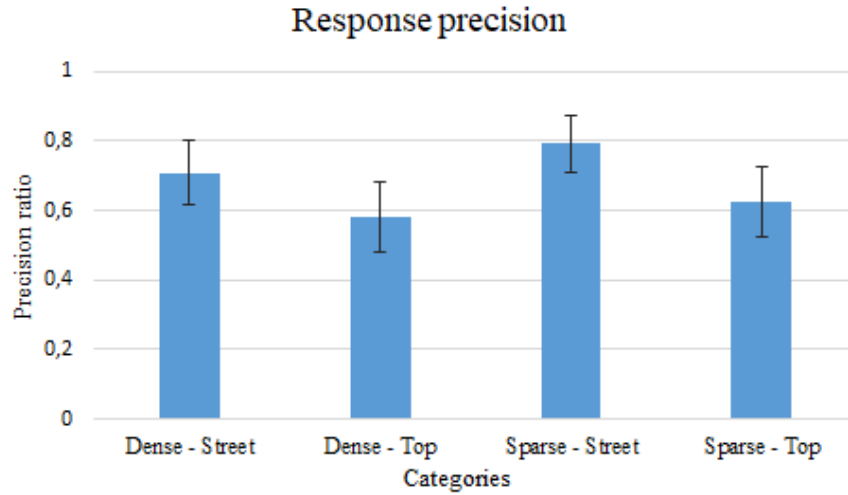


Figure 4.2: *Response precision with standard error. With permutations of variables density and perspective as categories. Precision ratio is calculated as number of correct responses to total number of responses.*

4.2 Test of hypothesis H_0

The degree of freedom is 1 for each variable respectively 91 for the residual or error. As mentioned before, a significance level of 5% was implemented. Hence, the f-critical value is 3.945694 for all variables. See table C.2 and C.3 in appendix C for a more thorough description of how ANOVA was implemented.

ANOVA results		
	F	F-critical
Density	10.22584	3.945694
Perspective	1.032969	3.945694
Both variables	0.270993	3.945694

Table 4.1: *Short version of table C.3 in appendix C. Only displaying relevant results for the subsection.*

Firstly, considering the variable density, $F(1,91) = 10.23 > 3.95$, hypothesis H_0 must be rejected in favour of H_1 with 95% confidence. Consequently, statistically density has a significant effect on users' perceived ability to identify low-resolution impostors.

Secondly, considering the perspective, $F(1,91) = 1.03 < 3.95$, hypothesis H_0 cannot be rejected in favour of H_1 with 95% confidence. Consequently, statistically perspective does not have a significant effect on users' perceived ability to identify low-resolution impostors.

Lastly, considering the combination of density and perspective, $F(1,91) = 0.27 < 3.95$, hypothesis H_0 cannot be rejected in favour of H_1 with 95% confidence. Consequently, statistically density and perspective interaction does not have a significant effect on users' perceived ability to identify low-resolution impostors.

4.3 Test of hypothesis H_1

In the previous section it was determined that H_0 holds with 95% confidence for both perspective and interaction between density and perspective. It remains, then, to be determined if H_1 must be rejected in favour of H_2 for variable density. If there is significance difference between the two. ¹

Confidence interval of precision		
	Lower bound	Upper bound
Sparse - Dense	1.68633	4.31367

Table 4.2: Difference in precision of spotting low-resolution impostors. Short version of table C.4 in appendix C. Only displaying relevant results for the subsection.

Being consistent with a 5% significance level it was concluded that there was a significant difference. Since, the confidence level is separated from 0. Therefore, with a 95% confidence H_1 can be rejected in favour of H_2 . Consequently, statistically, density has a significant effect on users' actual ability to identify low-resolution impostors.

¹The participants collected answers (right or wrong - 1 or 0) are binomial distributed which in turn could be approximated with a normal distribution. Which made it easier to deduce a confidence interval from a given level of significance. See equation C.1 and C.2 in the appendix.

5 Discussion

5.1 Analysis of results

The results show that, between the density of the crowd and the perspective from which it is viewed, the former variable has the most significant effect on viewers' sensitivity towards impostor resolution degradation. This holds both for viewers' perceived ability to distinguish which simulated crowd contains low-resolution impostors, as well as for their actual ability to do so.

A negative relation between crowd density and sensitivity towards impostor resolution degradation was found; regardless of from what perspective the crowd is seen, the mean precision rate at which viewers' identify the correct video clip was higher when crowds were sparse. This is likely due to sparse crowds allowing for easier and more thorough scrutinization of individual crowd members, which in turn makes successfully identifying 'off-looking', that is low-resolution, agents easier.

Also, a sparse crowd mitigates the general cluttering effect of viewing it at street-level. With relatively few agents in the scene, one can more easily scan individual agents at different distances from the camera, as indicated by figure 4.1. These results are consistent with findings on the negative relationship between crowd density and viewers' ability to identify crowd anomalies, as seen in section 2.1.1.

Interestingly, although dense crowds led to participants failing to identify which clip contains low-resolution impostors more frequently, the response time for these clips were shorter, as seen in figure 4.2. It would seem that high density crowds instilled in participants a false sense of confidence in their ability to identify the correct clip. It is not clear why this is, but one possible explanation is that of visual artefacts; with larger numbers of agents on the screen, cluttering and the number of visual artefacts increases. This, in turn, may lead to viewers, who may be simply looking for anything in the crowd that looks visually less appealing, to prematurely think they have found the clip containing low-quality agents.

Perspective, in contrast to the discarded corresponding hypothesis H_0 for density,

was not found to significantly affect participants' perceived ability to identify the correct clip.

Participant precision, on the other hand, was indeed affected by perspective, although not significantly. No matter the crowd density, the crowd containing low-resolution impostors were on average easier to identify at street-level. This is little surprising; in the study, when viewing the crowd at street level, agents are on average closer to the camera. This makes spotting low-resolution impostors easier, since they occupy more screen space. However, for the same reason, low-resolution impostors might also be occluded to a higher degree because of the impostors in front. If too much of the screen is occupied by high-resolution impostors at the front end of the crowd, this would counterbalance the effect described above. This could be why although study participants were more accurate in the street-level case, their response distance did not experience the same positive effect.

It is possible that the results were affected in favour of street-level clips by scene center bias (see Background section). As viewers' gaze tend to be drawn towards the center of the scene(s) in front of them, this could have different implications for street-level scenarios and top-down scenarios, since the contents of their center differs. Given a LOD threshold at x length units away from the far end of the crowd, low-resolution impostors (given high enough crowd density) can end up in the center of the scene of a street-level video clip. However, with a top-down view it might not even be possible for the low-resolution impostors to end up at the scene center, if the same LOD threshold is too close to the edge of the crowd.

5.2 Limitations

5.2.1 Split-screen stimuli

Study participants expressed a difficulty in that video clips were run in parallel. As participants were encouraged to identify which of the two clips contained low-resolution impostors, their gaze, naturally, switched back and forth between them. This may imply an added overall difficulty of scanning each crowd for whatever

they are encouraged to find (in this case, low-resolution impostors), and could possibly limit participants ability to identify the correct clip in the pair.

Similarly, because of the split-screen nature of the experiment's stimuli, a possible limitation is the fact that no two crowds shown simultaneously are the same crowd. Although both clips in a pair presents the same scenario (e.g. low density crowd in street view), the actual crowds simulated in each clip are different; they are from different runtimes of the simulation. It is possible, and not unreasonable to think, that impostors of lower resolution would have been easier to spot, had the crowds (apart from their impostors' image quality, that is) been identical. Some participants expressed that the crowds being different was distracting, although it is difficult to determine whether this actually affected their precision in identifying the correct clip. However, generating video clips with identical crowds where only the image quality of some impostors differ, would imply added significant difficulty as far as implementation is concerned.

It is worth mentioning that the possibly limiting factors described above hold (if they hold at all) for all video clips shown. It is within reason to think that the extent of their effect on the result is somewhat limited itself, since any pair of clips would be subject to the same viewer perception problems.

5.2.2 Gradual stimuli changes

It is possible that the rate at which the number of low-resolution impostors increases (i.e. LOD threshold distance decreasing by 5 percentage units every 5 seconds) is too subtle to notice. As a clip progresses, it may be that study participants become quickly accustomed to every LOD threshold, and so subtle gradual changes will not be identified, no matter how high the crowd fidelity at the start of a clip. Reasonably, suddenly moving the LOD threshold from the far end of a crowd to the front end (closest to the camera) would result in a degradation easier to spot than would the more gradual case.

5.2.3 Visual artefacts

The video clips generated for the evaluation featured quite simplistic crowds, in terms of agent homogeneity. Only three different models were used for simulating the crowds, which sometimes, albeit not unexpected, resulted in artefacts in the form of clusters of agents with the same model, such as four agents with the female model walking very closely to each other. While this is probably not directly related to viewers' ability to identify low-resolution impostors, it could be distracting for the viewing experience as a whole, especially in the street-level scenario, where this artefact becomes quite apparent as agents end up relatively close to the screen.

After the study had been conducted, the video clips used were deemed, both by participants and the authors, to be somewhat too dark. This may have affected participants' ability to differentiate the two crowds as low-resolution images become harder to tell apart from high-resolution ones.

Impostors are approximative 2D representations of 3D models, and as such, their visual fidelity is relatively low to begin with, at least in terms of animation quality, due to the limited amount of captured viewing angles from the 3D model. Although most participants did not realize that they were looking at 2D impostors rather than 3D models, it is possible that if the resolution of high-resolution impostors is too low, identifying the low-resolution impostors proves more difficult. This is especially hazardous if the 3D model on which the impostor is based is not very high-quality. As far as geometric detail level is concerned, the 3D models used for capturing impostor images (see Capturing section) in this implementation cannot be considered very high by today's standard in video games, film and other instances where virtual characters are found.

5.2.4 Impostor image quality and ambiguity

Related to the previous subsection's note on the possible difficulty in discerning high- from low-resolution impostors, there was a general confusion amongst study participants as to what it actually means for a crowd individual to be of 'low quality', or low resolution, which ultimately is what they were trying to identify.

To some viewers, it is not obvious how a high-resolution image differs from a low-resolution image, other than one being more visually appealing than the other. Study participants were not shown beforehand what a high- or low-resolution impostor looks like, and much less were they shown side by side. While this still seems reasonable from a methodological perspective, participants who, due to inexperience of the kind of virtual characters shown, were unsure of what ‘low-resolution agent’ means, might have a lesser chance of identifying the clips containing exactly that.

It is possible that showing study participants what kind of resolution degradation to expect would have mitigated this problem.

6 Conclusions

While perspective has some indicative implications for viewers' sensitivity towards impostor resolution degradation, ultimately it is crowd density which has a significant effect.

This effect has been measured and confirmed both in terms of viewers' perceived ability to identify crowds containing impostors of degrading resolution, and in terms of actual ability to do so. The relation is negative; higher density leads to a lower sensitivity, which is good news for applications where large, and thus probably dense, crowds are desirable. The result implies that as crowds become larger, we can get away with less detail without disrupting the viewing experience. This further strengthens impostor systems position as a ubiquitous LOD technique in entertainment and research, as demand for larger, denser and more realistic crowds grows.

7 References

- [1] Jacob von Eckermann. “How users differentiate impostors from real models”. BA thesis. 2008.
- [2] Daniel Thalmann. ”Crowd Simulation”. Encyclopedia of Computer Graphics and Games. 2016.
- [3] Francisco Arturo Rojas, Hyun Seung Yang, and Fernando M Tarnogol. “Safe Navigation of Pedestrians in Social Groups in a Virtual Urban Environment”. 2014.
- [4] Jack Shabo. “High Density Simulation of Crowds with Groups in Real-Time”. MA thesis. Kungliga Tekniska Högskolan. 2017.
- [5] Rahul Narain, Abhinav Golas, Sean Curtis, and Ming C. Lin. “Aggregate dynamics for dense crowd simulation”. 2009.
- [6] David Luebke, Martin Reddy, Jonathan Cohen, Amitabh Varshney, Benjamin Watson, and Robert Huebner. “Level of Detail for 3D Graphics”. 1st Edition. Elsevier, 2002.
- [7] Simon Dobbryn, Rachel McDonell, Michéal Larkin, Steven Collins, and Carol O’Sullivan. “Clone Attack! Perception of Crowd Variety”. 2008.
- [8] Keith O’Connor, Simon Dobbryn, John Hamill, and Carol O’Sullivan. “Geopostors: A real-time geometry / impostor crowd rendering system”. 2005.
- [9] Markus Bindemann. “Scene and screen center bias early eye movements in scene viewing”. 2010.

Appendices

Appendix - Contents

A Appendix - Instructions for Study	28
B Appendix - Questionnair for Study	29
C Appendix - Statistical calculations	30
References	26

A Appendix - Instructions for Study

Soon you will be shown eight short film clips. There will be two lanes with crowds walking towards you. In these clips the density of the crowd will vary as well as the perspective on the crowd.

Decided randomly one lane will contain high fidelity actors. While the other will contain low fidelity actors at a distance, as time passes the distance at which low fidelity actors is displayed will decrease.

The one difference between high and low fidelity actors is the resolution.

Goal: At each clip you shall determine which lane that contains low fidelity actors. As soon you are sure you shall give your answer.

Interaction: On the left and right hand side you will see a yellow button, degradation noticed here. When you are ready to give an answer click on the same side as you noticed the lesser fidelity.

After you locked in your answer a green button, start, will appear in the middle of the screen. When you are ready click on it to view the next movie.

B Appendix - Questionnaire for Study

Please fill in this short survey.

Age:					
Rate your vision, higher is better	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>
Mark your gender:	Female <input type="checkbox"/>		Male <input type="checkbox"/>	Other <input type="checkbox"/>	

This study intend to help answer a specific scientific question. The information given away in this user study is not connected to you personally, thus anonymous. This study in voluntary and you have the right to stop at any time. When the scientific question has been answered this information will be deleted.

I consent	<input type="checkbox"/>
-----------	--------------------------

C Appendix - Statistical calculations

Mean response distance						
Dense			Sparse			
	n	\bar{x}	σ	n	\bar{x}	σ
Top down	24	16.87500	3.867132	24	14.62500	4.105272
Street level	24	16.45833	4.173545	24	13.333333	4.310419

Table C.1: Mean response distance (and standard deviation) for each different stimuli.

Mean			
	Top - view	Street - view	Average
Dense	16.87500	16.45833	16.66667
Sparse	14.62500	13.33333	13.97917
Average	15.75000	14.89583	15.32292

Table C.2: Means calculated to perform the two-way ANOVA analysis of variance.

Two-way ANOVA					
	Sum squares	df	Mean square	F	F-critical
Density	173.3438	1	173.3438	10.22584	3.945694
Perspective	17.51042	1	17.51042	1.032969	3.945694
Both variables	4.593750	1	4.593750	0.270993	3.945694
Within (error)	1555.542	92	16.95154	-	-

Table C.3: Final calculation and results of the two-way ANOVA.

$$Bin(n, p) \sim N(np, \sqrt{np(1-p)}) \quad \text{if } np(1-p) > 10$$

Equation C.1: Approximation of binomial distribution as normal distribution.

Response precision					
	p	n	np	np(1-p)	$\sqrt{np(1-p)}$
Dense	0.625000	48	30	11.2500	3.35410
Sparse	0.687500	48	33	10.3125	3.21131

Table C.4: Calculations of the precision, which is binomial distributed.
Necessary calculations made to see if it suitable to approximate it as a normal distribution, see equation C.1.

$$\bar{X} - \bar{Y} \text{ is } N(\mu_1 - \mu_2, \sqrt{\sigma_1^2/n_1 - \sigma_2^2/n_2})$$

Equation C.1: Difference between two normal distributed samples with different variance.

Confidence interval					
	$\mu_1 - \mu_2$	σ_{1-2}	$\lambda_{\alpha/2}$	Lower bound	Upper bound
Sparse - Dense	3	0.670238	1.96	1.68633	4.31367

Table C.5: Final calculations and results of the confidence interval for the response precision. See equation C.2.

TRITA-EECS-EX-2019:379