Eye-catching Crowds: Saliency based Selective Variation

Rachel McDonnell[1] Michéal Larkin[1] Benjamín Hernández[2] Isaac Rudomin[2] Carol O'Sullivan[1]

[1]* Graphics, Vision and Visualisation Group, Trinity College Dublin
[2]† ITESM, Campus Estado de México.

Figure 1: *Variety types tested in the Selective Variation Experiment: (1) original character, (2) top texture variation, (3) face geometry variation, (4) face texture variation, (5 - 9) head accessories.*

Abstract

Populated virtual environments need to be simulated with as much variety as possible. By identifying the most salient parts of the scene and characters, available resources can be concentrated where they are needed most. In this paper, we investigate which body parts of virtual characters are most looked at in scenes containing duplicate characters or *clones*. Using an eye-tracking device, we recorded fixations on body parts while participants were asked to indicate whether clones were present or not. We found that the head and upper torso attract the majority of first fixations in a scene and are attended to most. This is true regardless of the orientation, presence or absence of motion, sex, age, size, and clothing style of the character. We developed a selective variation method to exploit this knowledge and perceptually validated our method. We found that selective colour variation is as effective at generating the illusion of variety as full colour variation. We then evaluated the effectiveness of four variation methods that varied only salient parts of the characters. We found that head accessories, top texture and face texture variation are all equally effective at creating variety, whereas facial geometry alterations are less so. Performance implications and guidelines are presented.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—[Rendering]

Keywords: eye-tracking, crowd rendering, virtual humans

*e-mail:{Rachel.McDonnell,Micheal.Larkin,Carol.OSullivan}@cs.tcd.ie
†e-mail:{hbenjamin, rudomin}@itesm.mx

1 Introduction

Simulating heterogeneous crowds is a challenging task. The assets required to vary humans, such as textures and accessories, can be expensive to purchase, time-consuming to create and require extensive memory and computing resources. In this paper, we address these issues by developing a selective variation method for virtual humans. The method is based on the results of our eye-tracking experiments, where we identify the features of characters that are most attended to. More specifically, we consider: which body parts are most important to vary? Will this differ based on orientation and the presence or absence of motion? Do different characters require different body part variations? Will selective variation based on human eye-movement patterns be as effective at disguising clones as full variation?

In our body-part saliency experiments, we investigate whether people focus visual attention on certain parts of virtual characters more than others, when asked to assess the variety of a crowd. We tested a range of different human characters and found that attention is focused almost exclusively on the head and upper body, regardless of the character. Based on this information, we developed a selective variation technique, which varied only salient body parts, and validated our approach. In a final set of experiments, we tested the effectiveness of several other selective variation techniques: facial geometry variation, facial texture variation, garment texture variation, and accessories.

We apply our method to render real-time crowds, but our results should also be useful in all scenarios where varied crowds are needed and resources need to be optimised. In Section 5, we provide a performance evaluation to indicate the amount of texture memory, time, and computation that could be saved.

2 Background

The simplest and most popular method for increasing the variety of virtual humans in a crowd is hardware accelerated per body part colour modulation [Tecchia et al. 2002; Gosselin et al. 2005; de Heras Ciechomski et al. 2005a; Maïm et al. 2007]. The body of the character is segmented into different parts, usually by encoding the different regions in an alpha map, which is associated with the diffuse texture of the character. The final colour of each body part is a modulation of its texture colour. A colour can be randomly selected, but this often results in undesirable colour combinations. A more common technique is to manually create a set of 'outfit maps' for each character. These outfit maps contain the HSV offsets for each body part. Since they are manually chosen, the resulting combinations are natural. However, this can be a time-consuming task, particularly if the number of regions is high.

More recently, accessories and texture modulation have been used to modify the appearance of cloned characters so that they appear different [Thalmann and Musse 2007]. Geometry variation is another solution, where the character meshes are broken into multiple pieces and alternative meshes are provided for variation [Dudash 2007].

These techniques are effective but require increasing amounts of texture memory and computation time as the number of varied or accessorised body parts grows. Knowing the parts of the body that are most focused on when viewing crowds would be very useful, as texture memory consumption and computation time could be reduced by varying only those parts of the body that receive the most attention. In [McDonnell et al. 2008], we studied the perceptual impact of cloning characters and their motions. We found that appearance variation is most important for creating the illusion of a varied, heterogeneous crowd.

Yarbus [1967] showed that, even under static viewing conditions, not every object in the field of view will capture visual attention. Researchers have tried to exploit this fact, by predicting the focus of attention and decreasing the rendering quality or level of detail in the less important areas. Itti et al. [1998] provided a computational model of visual attention that identifies the possible areas of visual attention. Yee et al. [2001] created a hybrid vision model, based on [Itti et al. 1998] combined with spatiotemporal sensitivity [Daly 1998], in order to significantly accelerate global illumination calculations in pre-rendered animations. More recently, Peters and Itti [2006] developed simple neurally-inspired algorithmic methods to predict where humans look while playing video games.

Research on mesh simplification has also benefitted from exploiting visual attention. Lee et al. [2005] developed a mesh saliency operator, inspired by human perception, which resulted in more visually pleasing results in the processing and viewing of meshes, compared with using geometric measures of shape.

It has been shown that visual attention is largely controlled by task [Yarbus 1967]. In a series of experiments, Cater et al. [2003] proved that it is possible to render scene objects not related to the task at a lower resolution, without the viewer noticing a reduction in quality. Howlett et al. [2005] used eye-tracking data to examine the role of feature saliency in model simplification and provided guidelines for perceptually guided simplification. They found that for natural familiar objects, features like the head, eyes and mouth were viewed more than others.

Research on the perception of human faces suggests that faces are processed in a different cognitive manner than other objects. Studies have indicated that humans possess a face-specific recognition system [Farah 1992]. Furthermore, Hochstein et al. [2004] found that faces popped out from a background of varied photograph distracters. However, other evidence suggests that faces do not 'pop-out' during visual search and are not accessed pre-attentively in parallel [Nothdurft 1993]. Sinha et al. [2006] investigated factors that influence facial recognition. They found that colour, pigmentation and shape were important cues for identifying individuals. We also exploit such alterations to try to disguise clones.

3 Body-part Saliency Experiments

Perception research has shown that the eye tends to fixate and saccade between regions of interest repeatedly [Yarbus 1967; Noton and Stark 1971]. Fixation data can be used to determine saliency, since fixations indicate the viewer's spatial focus of attention over time. In this first set of experiments, we wish to determine the most salient parts of virtual characters. In the *Appearance* experiment we determine the most salient areas of characters in the absence of motion. We also investigate whether or not the salient areas vary based on the orientation of the character, or the model used. In the *Appearance and Motion* experiment we determine the salient areas when motion and appearance are combined. In the *Motion* experiment, we investigate which areas are most salient when motion alone is present. We found that participants fixated on different body parts for appearance alone and motion alone. When motion and appearance were combined, their fixations were influenced entirely by the appearance rather than the motion.

3.1 Models, Motions and Framework

A set of 5 female and 5 male models was used, representing typical pedestrian types; aged young and middle-aged and wearing mainly casual attire. Each model had a single texture map which included all of the textures for their skin, hair and clothing (photographed from real people). We manually created an alpha map which encoded each region with a unique greyscale value (as in [de Heras Ciechomski et al. 2005b]). These values were then used to index an outfit map at runtime. Since we wished to examine characters with all regions varied in this experiment, we manually created 32 unique outfits for each model with varying hair, skin, clothing and shoe colour. For the experiment with motion alone, we used a neutral mannequin model.

We also captured the walk motions of five male and five female volunteers. Each of these walks was applied to the mannequin model for the Motion Experiments, and matched with an appropriate character model.

Figure 2: *Mesh split into 14 different regions.*

Since we wished to compare fixations on body parts across all of the models, we chose 14 common areas for each character mesh, rather than using the same regions as encoded in the alpha map. Figure 2 illustrates the 14 body parts that were used: head, upper and lower torso, upper and lower left and right arm, pelvis, upper and lower left and right leg and left and right foot.

Eight characters were displayed on a grey ground plane with no shadows. They were arranged so that they appeared as large as possible on-screen, with no occlusions (Figure 3). The experimental crowd system was developed using an open-source renderer and the HSV color modulation was implemented as a High Level Shader Language (HLSL) program. The experiments were viewed on a wide-screen 24 inch LCD monitor with a resolution of 1920 * 1200 pixels.

Figure 3: *Example scene from front orientation condition.*

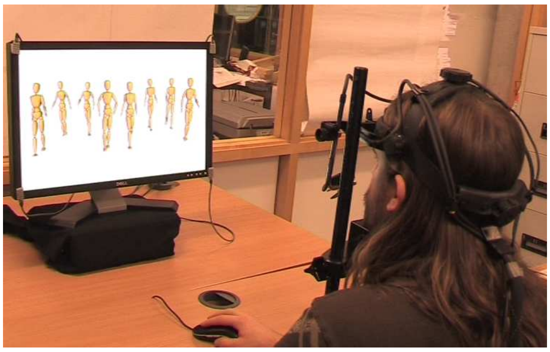

Figure 4: *Participant in the motion saliency experiment.*

We used an SMI EyelinkII eye-tracking device to record fixations. The eye-tracker is accurate to within 0.5 degrees of visual angle. The eyes were both tracked at 500hz and we recorded the *xy* screen coordinates of fixations. A ray was then cast from the fixation point into the scene to determine the body part of the character that was being attended to. To account for slight inaccuracies in the fixation point, we did a nearest-neighbour search on the 100 pixels surrounding the fixation point, to determine the closest body part.

Participants sat at a constant distance of 60cm from the display. This distance was maintained by instructing them to place their head on a chin-rest for the entire experiment (Figure 4). The chin-rest was also useful for minimising head drift, which increased the accuracy of the eye-tracking. When a participant fixated on one of the 14 body regions of one of the character meshes, the duration of the fixation was recorded and stored for that region. We also recorded the number of individual fixations on each region. Characters were displayed on-screen such that the visual angle subtending the smallest body region was within the range of accuracy for the eye-tracking device used.

3.2 Appearance Experiment

Fourteen volunteers (8M-6F) participated in this experiment and were given a book voucher as a token of appreciation (as in all our subsequent experiments). All were naïve as to the purpose of the experiment and from different educational backgrounds. During the experiment, participants wore an eye-tracking device in order to record the necessary data.

They were first informed as to what a colour modulated clone was (using an example of a model not used in the experiment). Then, two rows of four models were shown on-screen. The minimum number of colour clones dispersed among the crowd was 0 (i.e., all different) and the maximum was 5. Models were static and oriented forwards in a neutral pose that allowed all body parts to be fixated on. For each trial, participants were asked to answer the question: "Does this scene contain clones? Yes or No". As discussed in Section 2, task can influence visual attention. We chose the task of identifying clones as it is the areas that are focused upon when a clone is spotted that crowd developers will be most interested in disguising.

Participants indicated their decision by clicking the left or right mouse button, whereupon the screen automatically changed to perform head drift correction between each trial. We displayed 15 trials where clones were present (5 number of clones * 3 repetitions). Since the task was to indicate the presence or absence of clones, we added a further 15 trials where no clones were present (all 8 characters were different) in order to ensure that a bias towards answering "yes" did not occur. The 30 trials were viewed in random order by each participant. All clones were colour modulated, and the cloned character was chosen randomly at each trial. All other non-cloned characters in the scene were different from each other.

Figure 5: *Example scene from side orientation condition.*

Figure 6: *Example scene from back orientation condition.*

In order to test the effect of orientation, three blocks of conditions were tested. In the first block, all characters were oriented facing forwards (Figure 3). In the second block, they were orientated facing to the right, and in the third block they faced backwards (Figures 5 and 6). Participants viewed the blocks in random order, to avoid ordering effects.

As previously discussed, we hypothesised that people would focus on the heads of characters when the their faces were visible (orientated facing forward). When oriented away from the camera, it seemed likely that the head would not be focused on as much. Furthermore, we predicted that the torsos of the characters with logos on their shirts and the legs of characters wearing shorts or skirts would also be salient.

3.2.1 Results

Unless otherwise stated, for all of our experiments we performed a two-way repeated measures ANalysis Of VAriance (ANOVA) for main effects with post-hoc analysis using Newman-Keuls comparisons of means.

Participant Accuracy

First, we analysed participant performance in the task of detecting the presence or absence of clones. We averaged participants' results over the number of repetitions, for every number of clones. We then performed a two factor ANOVA on the data, where the conditions tested were *orientation* (3) and *number of clones* (5). We found no effect of orientation, which implies that participants were equally accurate regardless of the orientation of the character. We found a main effect of the number of clones ($F_{4,52} = 23.96, p < 0.00001$) (Figure 7). This effect was due to the fact that for the case where one clone was present, participants answered incorrectly significantly more often than for all other numbers of clones (p <0.0005 in all cases). Furthermore, in the control case (where there were no clones on-screen) there was a false positive rate of 10%. False positives probably occurred due to participants mistaking similar characters for clones.

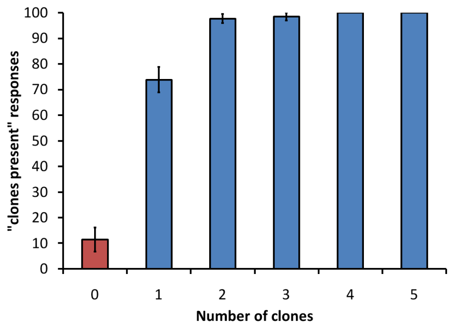

Figure 7: *Percentage of "clones present" responses, based on the number of clones on-screen.*

Fixation Duration

One metric that can be used to analyse eye-fixation data is the average duration of all fixations on a region [Henderson and Hollingworth 1998]. We first looked at the average fixation duration on each of the body parts for each orientation, averaged over the number of clones on-screen. A two factor ANOVA was performed, where the factors were *orientation* (3) and *body part* (14). A main effect of body part was found ($F_{13,169} = 30.26, p < 0.000001$) (Figure 8). Post-hoc analysis showed that *the head was fixated on significantly more often than any other body part*. The upper torso was the next most significantly fixated on. All other body parts were fixated on equally often, except for the lower torso which was fixated on more often than the right foot (p <0.05 in all the above cases).

There was no main effect of orientation. However, a body part-orientation interaction did occur ($F_{26,338} = 9.01, p < 0.000001$). Post-hoc analysis showed that this effect was due to differences in the length of time spent fixating on the head and upper torso for the three orientations (Figure 9). We also found that: the head was

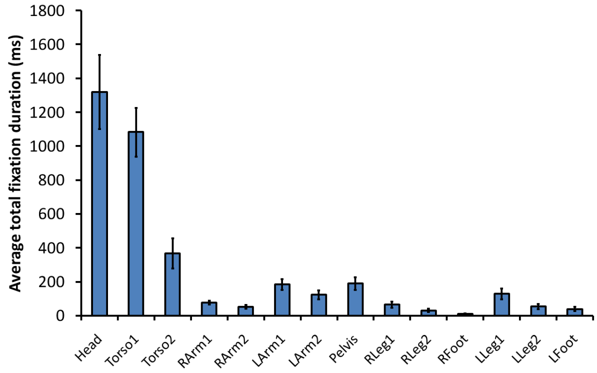

Figure 8: *Fixation duration main effect of body part over all orientations. Labels: L, R, 1 and 2 represent Left, Right, Upper and Lower respectively.*

fixated on more often in the Side than in the Front or Back orientations; the Upper Torso was fixated on more often in the Front and Back than in the Side orientation; and the head was fixated on more often in the Front than the Back orientation (p <0.00005 in all cases).

We also performed a two factor ANOVA on the fixation duration data, where the factors were *number of clones* (6) and *body part* (14). We found that fixation patterns on body parts did not change with the number of clones. However, as expected, the length of time they were focused on decreased with increasing numbers of clones.

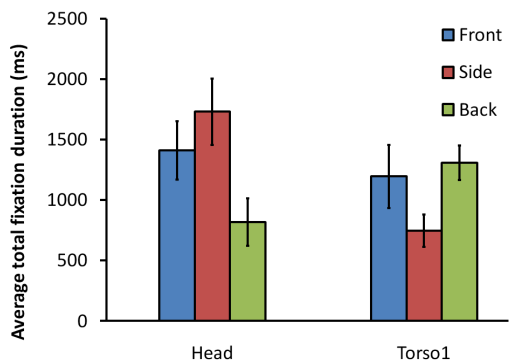

Figure 9: *Interaction between orientation and body part.*

Number of Fixations

We also recorded the number of times that participants fixated on different body parts [Laarni et al. 2003]. Based on this data and inspection of the eye-movements recordings, we found that the total duration times discussed above were made up of multiple fixations. This implies that participants did not stare at the salient parts for long periods of time. Rather, they fixated for short periods and performed saccades regularly between characters to scan the crowd for clones.

Percentage fixations

Before testing whether different body-parts were focused on for different characters, we first normalised duration values (because some characters may have been displayed more often than others due to random selection). We thus calculated percentage fixation durations for each body part for each character.

We first collapsed the values over orientation and then performed a two factor ANOVA where the conditions were *body part* (14) and *character* (10). As before, the same main effect of body part was found ($F_{13,169} = 23.36, p < 0.00001$). There was also an interaction between character and body part ($F_{117,1521} = 2.13, p < 0.00001$). From observing the graphs it was found that the interaction was mainly due to either the head or the torso being viewed more often than the other, but no particular reason for this difference was obvious. No other body part was focused on significantly more than the head or the upper torso. This result implies that *the bottom half of the body is almost never focused on, regardless of whether the character is male/female or wearing trousers, skirts or shorts*.

First Fixation Point

The first fixation point in a scene can be a good metric for attention capture [Henderson 1992]. We performed a two factor ANOVA on this data where the conditions were *orientation* (3) and *body part* (14). A main effect of body part was found ($F_{13,169} = 28.56, p < 0.000001$). Post-hoc analysis showed that the first fixation was most often on the head (Figure 10), followed by the upper torso, then the lower torso ($p < 0.0005$ in all cases). The head, upper torso and lower torso were fixated on first an average of 36%, 25% and 10% of the time. For all other body parts, the percentage of first fixations was between 0% and 6%. These findings are consistent with other studies discussed in Section 2.

A main effect of orientation was not found, but an interaction between body part and orientation was ($F_{26,337} = 7.3, p < 0.000001$). Post-hoc analysis showed that this interaction was due to the same differences in Head and Upper Torso as above ($p < 0.05$ in all cases). Additionally, it was found that the lower torso captured fixations more often for the Back view than for the Side or Front views ($p < 0.05$ in both cases).

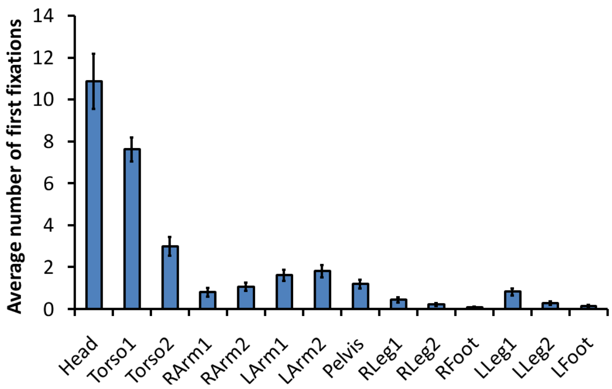

Figure 10: *First fixation main effect of body part over all orientations.*

3.3 Appearance and Motion Experiment

Characters in crowd simulations often have cloned motions as well as cloned appearances. The most common approach is to assign each template character its own characteristic motion. In this experiment, we wished to determine if the salient regions of characters with cloned appearance were affected by their cloned motions.

We know from previous point light experiments that motion is a strong cue for individuality [Cutting and Kozlowski 1977; Beardsworth and Buckner 1981]. It is possible to identify an individual based on their walking motion alone. Therefore, we hypothesised that adding a characteristic motion to our characters could possibly change the fixation patterns of participants.

Fifteen new participants (10M-5F) took part in this experiment. All participants were naïve to the purpose of the experiment and from different educational backgrounds. Each of the ten character models was matched with one of the ten walk motions captured. A clone in this experiment had cloned motion and appearance. All characters faced forward and cloned motions were played in-step. All other experiment conditions were as in the previous experiment.

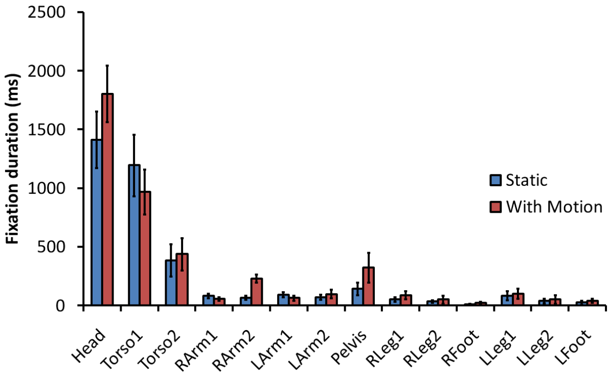

Figure 11: *No main effect of the presence or absence of motion.*

3.3.1 Results

A single factor ANOVA was performed on the results which showed a main effect of body part ($F_{13,182} = 33.59, p < 0.000001$). Post hoc analysis showed that this was due to the head being fixated on significantly more often than any other body part, followed by the upper torso ($p <$0.00005 in both cases).

In order to gain insight into whether or not motion had an effect on what was being fixated on, a between-groups two factor ANOVA was performed comparing this data to that of the front orientated static condition. The factors were *body part* (14) and *motion* (2). We found no significant effect of the presence or absence of motion and no interaction (Figure 11). This implies that *the appearance of the characters completely dominated, and cloned motion had no effect on where participants fixated*.

3.4 Motion Experiment

In our next experiment, we isolated the effect of cloned motion in order to determine whether the same body parts were fixated upon as in the previous cases.

Fifteen new naïve participants (9M-6F), from different educational backgrounds, wore an eye-tracking device in order to record their fixations. A neutral mannequin model was used in this experiment so that participants could focus on the motion alone. Two rows of four mannequins were displayed on-screen (Figure 12). The experiment procedure was identical to that of the previous experiments, except that motions were cloned rather than models. We did not test orientation in this experiment as preliminary trials with the mannequins facing to the side or backwards proved too difficult. Therefore, the mannequins always faced forwards, and clones were displayed while walking in-step. Participants were informed as to what a motion clone was, using an example of a walk motion not present in the experiment.

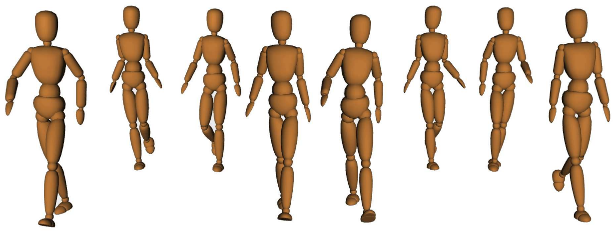

Figure 12: *Example scene from the Motion Experiment.*

3.4.1 Results

Fixation Duration

A two factor ANOVA was performed on average fixation duration data, where the conditions were *body part* (14) and *number of clones* (5). A main effect of body part was found ($F_{13,195} = 10.44, p < 0.000001$). This effect was due to the pelvis and lower torso being fixated on equally often and significantly more than on any other body part (p <0.0005) (Figure 13). A main effect of clone number was also found ($F_{5,75} = 13.55, p < 0.000001$). As before, this was due to an increase in time spent fixating on a scene, the fewer clones there were on-screen. An interaction also occurred ($F_{65,975} = 2.62, p < 0.000001$). The number of clones shown had more of an effect on fixation duration for the body parts that had most fixations overall (pelvis and lower torso).

These results indicate that when the appearance of the characters is the same, but their motions are cloned, the head and upper torso are not fixated on at all, but rather the middle region of the body. This is interesting considering recent research that indicates that hip sway is an important factor for detecting the differences between male and female gaits [Johnson and Tassinary 2005]. With this in mind, we averaged the fixation duration values over the sex of the walker, in order to determine if participants fixated on different areas for male and female walkers. A two factor ANOVA was performed on the data where the conditions were *body part* (14) and *sex of walker* (2). As before, a main effect of body part was found. However, we found no effect of the sex of the walker and no interaction, meaning that the pelvis and lower torso were fixated on equally often, regardless of the sex of the walker. This indicates that *the sway of the character could be an important factor to vary in order to create variety in walking motions*, particularly for scenes where appearance variation is necessarily low (e.g., a crowd of business people wearing suits).

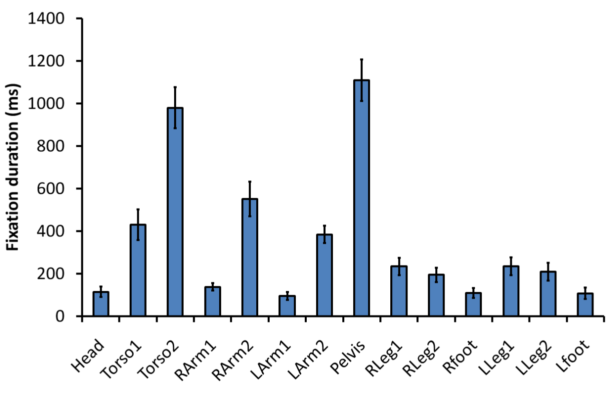

Figure 13: *Main effect of body part in Motion Experiment.*

3.5 Validation

The results of this set of Saliency experiments showed that the head and torso are the most salient features of virtual character models. Using this fixation information, we propose a method for selective body-part variation, where only the colour of salient body parts are varied. To test the feasibility of this approach, we compared clone detection performance on crowds exhibiting different types of colour variation.

3.5.1 Experiment Procedure

Ten naïve volunteers (8M-2F) from different educational backgrounds, of whom none had participated in the previous study, took part in this experiment. A full set of twenty different template models (including the ten from the previous experiment) was used. Exactly 20 characters, some of whom were clones, were displayed in four rows of five characters, each placed so that they were optimally visible. In [McDonnell et al. 2008], we used reaction times for clicking on pairs of clones as a measure of how effectively clones were disguised. We use this same metric in order to determine the effectiveness of varying only certain body parts. Participants were asked to click on the first pair of clones they saw in the scene, as quickly as possible. Reaction times were recorded for each trial.

Randomly dispersed amongst the twenty characters in the crowd were 2, 4, 6 or 8 clones. The model chosen to be cloned was random for every trial, and all other characters in the scene were unique, with randomly chosen outfits. The condition that we wished to examine in this experiment was *variation type* (4). The four variation types that we used were: *no variation* (which meant that clones were identical), *full* colour variation, *selective* salient body-part colour variation and *inverse-selective* body-part colour variation (which meant that salient parts were not varied, but all other parts were). In order to maintain the natural appearance of each character, variation was based on the alpha regions that incorporated the head and upper torso areas.

Four repetitions of each condition were displayed, resulting in a total of 64 randomly viewed trials: 4 numbers of clones * 4 variation types * 4 repetitions. We hypothesised that the exact clones and inverse-selective clones would be spotted the quickest, with selective salient and full variation clones being more difficult to detect.

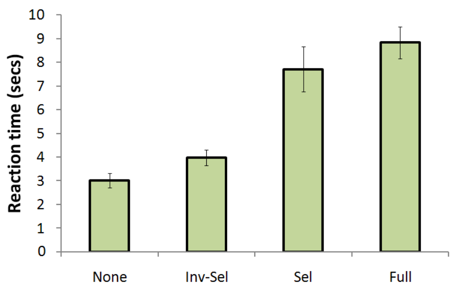

Figure 14: *Main effect of variation type.*

3.5.2 Results

The average reaction times for each of the variation types were found for each participant (averaged over the number of clones). A single factor ANOVA showed a main effect of variation type ($F_{3,27} = 29.65, p < 0.00001$) (Figure 14). Post-hoc analysis showed that there was no significant difference between exact and inverse-selective variations, which took on average 3 and 3.9 seconds for participants to detect clones. Furthermore, there was no significant difference between selective and full variations, which took on average 7.7 and 8.8 seconds for participants to identify clones.

Significant differences were found between selective and both inversive-selective and exact, and between full and both inversive-selective and exact ($p < 0.0005$ in all cases). This implies that, even though the small difference between full and selective variation may become significant with more participants, it is clear that *only varying the colour of salient body parts is almost as effective at disguising clones as using full colour variation*. Also interesting to note is how the inverse-selective method was no more effective than exact. This implies that *varying the lower half of the body does not produce additional variety*.

4 Selective Variation Experiments

In the Body-Part Saliency experiments, we established that some body parts attract more attention than others. Varying the colour of these parts only is as effective as varying the full body. We now evaluate several popular methods for disguising the salient regions of virtual humans. In order to provide useful developer guidelines, we analyse the effectiveness and practicality of these techniques in terms of user perception and rendering costs.

Figure 15: *Facial texture variation.*

4.1 Selective Variation Techniques

In order to disguise the salient regions of the characters, we used several different variation techniques. The torso was disguised using *top garment texture* variation. This involved storing three different top textures for each character, created from photographs of real clothing. In order to disguise the face, we tested three different techniques. The first was *facial texture* variation, which involved converting each model's face textures into a generic texture space, where all facial features of each model were in full correspondence. Then, a number of images that could be blended in realtime with these base textures were created, with make-up for the females and beards for the males (Figure 15).

The second technique was *facial geometry* variation. Two displacement maps were generated for each character. Different values were used to alter the characters' cheekbones, eyebrows, nose, mouth and chin. We hand-picked the maps in order that the face appeared as different as possible to the original, whilst retaining a natural appearance. The final technique tested was *accessories*, which have been used in the past to vary the appearance of characters [Thalmann et al. 2007]. Two different pairs of glasses were stored for each character, along with a hat and hairstyle. In order to maintain the natural appearance of our characters, we chose accessories that would look plausible on pedestrians.

We hypothesised that accessories would be most effective at disguising clones, in particular the hats and hairstyles since they change the silhouette of the characters' heads. We also expected that the top garment texture variation would be more effective than the face texture or geometry variation since the garment occupied a much larger area than the face.

4.2 Testing Variation Methods

Twenty naïve participants volunteered for this experiment (12M-8F). Ten characters were displayed on the screen in perspective. A single randomly chosen character in the crowd was cloned, and participants were asked to find them as quickly as possible. We recorded the reaction time from the start of the trial until the first character of the correct pair was clicked. False positives were also recorded. We ensured that there were always five male and five female characters on-screen so that a larger number present of one sex could not be used as an indicator of the sex of the cloned character. The randomly chosen character was always displayed as the original and its clone was disguised using the condition being tested. All other characters were given a 50% chance of being displayed in their original state or of being varied using the condition being tested.

This experiment was split into 6 blocks, each consisting of 15 repetitions of the condition being tested. Block 1 tested the exact clone condition, where the cloned character was identical to the original. This condition was included to allow participants to become familiar with the characters and the experiment setup. This block was always shown to participants first as a training block, and all other blocks were shown in random order, to avoid ordering effects. Block 2 tested the colour clone condition, where all characters were colour modulated, including the chosen character and its clone. We included this condition as the baseline or control case.

All characters in blocks 3 to 5 used colour modulation, since this technique is the simplest to implement and commonly used in crowd systems. Block 3 tested facial texture variation, where the cloned character was disguised using a blended beard or make-up. Facial geometry variation was used to disguise the clones in block 4, while top garment texture variation was used in block 5. In block 6 we tested 3 repetitions of 5 accessory types: glasses, hat, hairstyle, glasses and a hat, and glasses and a hairstyle.

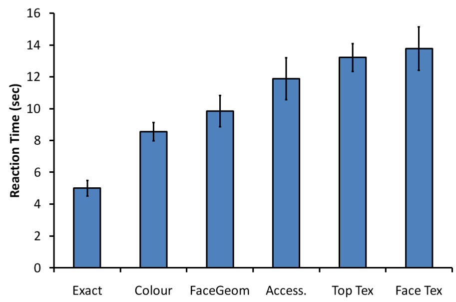

Figure 16: *Main effect of variation type.*

Results: We gathered the reaction times from each of the participants for each of the variation types, and averaged over the 15 repetitions. A two factor ANOVA was performed on the results where the conditions were *body part* (14) and *variation type* (6). A main effect of variation type was found ($F_{5,75} = 17.49, p < 0.000001$), as shown in Figure 16. As expected, the exact clones were easier to spot than any other variation type, followed by colour variation clones (p <0.005 in all cases). Facial geometry was the next easiest to detect (p <0.01 w.r.t. facial texture and top texture). Finally, *accessories, facial texture and top texture variation were found to be equally effective at disguising clones*.

In order to determine if the different accessory types had an effect, we ran the accessories block on a further 9 participants so that this condition was viewed by 29 participants in total (19M-10F). We averaged over the 3 repetitions and performed a single factor ANOVA on the results, shown in Figure 17. A main effect of accessory type was found ($F_{4,112} = 3.69, p < 0.01$). Post hoc analysis showed that *the addition of glasses along with a change of hairstyle was more effective at disguising clones than any other accessory type* (p <0.05 in all cases).

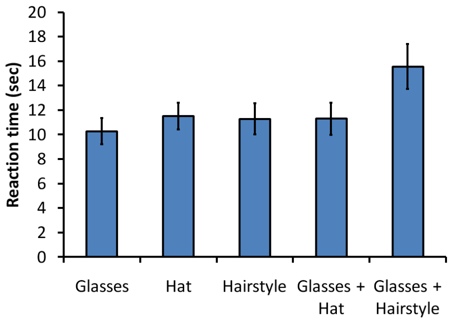

Figure 17: *Main effect of accessory type.*

We summed the number of false positives that occurred for each experiment block, for each participant. A single factor ANOVA showed a main effect of variation type ($F_{5,95} = 5.64, p < 0.0005$), where exact, colour, and facial geometry clones had fewer false positives than all other variation types (p <0.05 in all cases). Figure 18 shows that female characters were rarely confused with male characters, and vice versa. Furthermore, some male characters were more often confused than others (Figure 19).

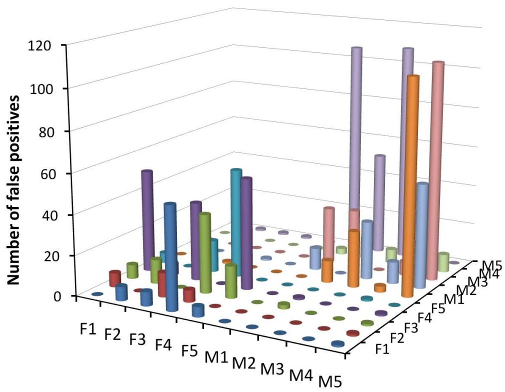

Figure 18: *False positive confusion matrix. F1 - F5 are female models and M1 - M5 are male models. The blank areas of the graph show that males were never confused with females and vice versa. The data was accumulated from all trials for all participants (1500 trials in total).*

4.3 Testing Orientation

Face texture variation proved as effective as top garment variation and accessories from the front viewpoint. However, from certain orientations (such as the back) this type of variation would not be effective at all since it would not be seen. Since orientation affected the duration of fixations on head and torso differently in the Saliency experiments, we ran a further 5 experiment blocks to test this effect. Fourteen naïve volunteers (5F-9M) viewed the 5 blocks

Figure 19: *The character on the left was most often confused with the two characters to his right.*

in random order to avoid ordering effects. Blocks 1 and 2 tested accessories from the side and back respectively, blocks 3 and 4 tested top garment texture variation, while block 5 tested facial texture variation from the side only.

Results: For the back orientation, we found a main effect of variation type ($F_{1,13} = 14.65, p < 0.005$), due to the fact that top garment variation made clone detection significantly more difficult than variation using head accessories (Figure 20). This is consistent with the results of the Body-Part Saliency experiments, where it was found that the torso was focused on more often than the head for the back orientation. There were no interesting effects for the side orientation.

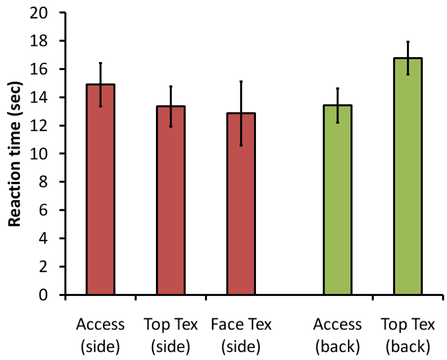

Figure 20: *Back and front orientation reaction times.*

4.4 Testing Crowd Variation

In this experiment, we tested the effectiveness of selective variation techniques on a large crowd, using a more general task involving the perception of variety.

Twenty-five naïve participants volunteered for this experiment (11M-14F), and viewed a series of 10-second movies depicting large crowds. They were informed that a set of character models had been cloned to create a crowd of 360 characters, and that different techniques had been used to disguise these clones. However, they were not shown examples as we wanted their unbiased reaction to the level of variety present. The task was to indicate on a 5-point scale the level of clones that they thought were present, ranging from *hardly any* to *lots* of clones. We tested whether adding selective texture variation, head accessories, and increasing numbers of template models enhanced variety. Twenty-seven movies were shown to each participant in random order: 3 variation types (colour, colour + selective texture, colour + selective texture + accessories) * 3 templates numbers (2, 4, 6) * 3 repetitions.

Results: We gathered the results from each of the participants and averaged over the 3 repetitions. A two factor ANOVA was performed on the results where the conditions were *variation type* (3) and *number of templates* (3). A main effect of the number of templates was found ($F_{2,46} = 31.57, p < 0.00001$). As expected, crowds using increasing numbers of template models were judged as having fewer clones (p <0.0001). There was also a main effect of variation type ($F_{2,46} = 35.67, p < 0.00001$). Post hoc analysis showed that *incrementally adding selective variation results in fewer clones being detected in a large crowd*. Post hoc analysis showed that crowds using only 2 templates with colour, texture and accessory variety were considered to be equally heterogeneous to those using 6 templates with colour variety only (Figures 21 and 22). This shows that *using selective variation techniques can significantly reduce the number of template models required to create a varied crowd.*

Figure 21: *(top) crowd using 6 template models with colour variation only, (bottom) crowd using 2 template models with colour, texture and accessory variation.*

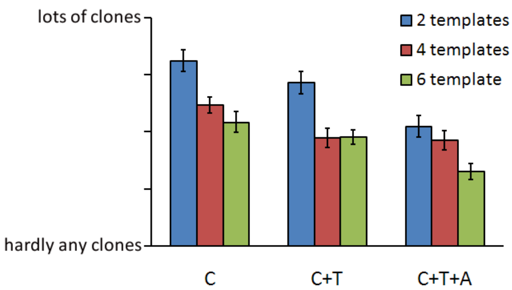

Figure 22: *Results from Crowd Experiment. C = colour variety alone, C+T = colour + texture variety, C+T+A = colour + texture + accessories.*

5 Performance, Conclusions & Future Work

For cloned characters using colour modulation alone, the advantage of using selective variation is that it takes less time to create the HSV outfit maps and the encoding of the alpha channel regions. Per model, it takes on average 30 minutes to outline the alpha channel regions of the image. This time would be reduced significantly, if only the top half of the character needed colour modulation. Furthermore, it takes up to 5 minutes per outfit map to create natural looking HSV offsets, which could also be reduced.

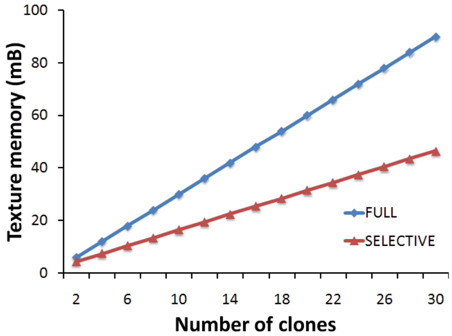

Figure 23: *Texture memory consumption reduction for selective texture variation based on one character cloned.*

Selective body texture variation proved effective at disguising clones. The advantage over using full texture variety is the reduction in texture memory and the cost of commissioning textures. Currently, 3MB of memory are needed to store a 1024 * 1024 RGB image. A single 512 * 1024 RGB image could be stored for the lower half of the character, per template model. Multiple 512 * 1024 RGB images could then be stored to create a range of different textures for the top half of the character. Figure 23 shows the amount of texture memory saved by using selective texture modulation rather than full texture modulation. For large crowds with many cloned templates, this represents significant savings.

Facial texture variation proved as effective at disguising clones as body texture or head accessories. It is also the least resource intensive and the least computationally expensive of the three methods. If all faces are in the same texture space (as with our models), only a single set of facial textures is needed, which can be blended at runtime with minimal computation. Alternatively, the facial textures could be pre-blended into the original texture with an extra face texture stored for each variation required.

Head accessories proved as effective at disguising clones as the face and top texture variations. Although accessories add to the realism of the crowd, there is a computational overhead due to the additional polygon count and animation updates. We therefore recommend that they be used sparingly for added realism, but that face and body texture variation be used more extensively to create the impression of a varied, heterogeneous crowd.

As reported by Yarbus et al. [1967], task has an effect on fixation patterns. In our Body-part Saliency experiments, we used the task of spotting clones, which is arguably the worst case scenario for a crowd system. Clone detection would clearly be lower when participants were not actually looking for clones, as their fixations would be more concentrated on task related objects. However, it may be the case that more unusual characters or animations would attract attention in a different way (e.g., clones of a limping character could draw attention to the legs). In future work, we would like to investigate attention effects with a wider variety of characters and animations.

Interesting to note is how facial geometry alterations were not found to be very effective at disguising clones in our experiments. However, Sinha et al. [2006] did find that shape alterations hindered face recognition. It is possible that either the distance at which the characters were displayed had an effect, or perhaps changing the overall shape of the head may have been more effective than altering the shape of the facial features alone. Future work will investigate these possibilities.

Acknowledgements

This research was sponsored by Science Foundation Ireland as part of the Metropolis project.

References

BEARDSWORTH, T., AND BUCKNER, T. 1981. The ability to recognize oneself from a video recording of ones movements without seeing ones body. *Bulletin of the Psychonomic Society 18*, 1, 19–22.

CATER, K., CHALMERS, A., AND WARD, G. 2003. Detail to attention: Exploiting visual tasks for selective rendering. In *Proceedings of the Eurographics Symposium on Rendering*, 270–280.

CUTTING, J., AND KOZLOWSKI, L. 1977. Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society 9*, 5, 353–356.

DALY, S. J. 1998. Engineering observations from spatiovelocity and spatiotemporal visual models. *Proceedings of SPIE 3299*, 180–191.

DE HERAS CIECHOMSKI, P., SCHERTENLEIB, S., MAÏM, J., AND THALMANN, D. 2005. Real-time shader rendering for crowds in virtual heritage. In *Proceedings of the 6th international Symposium on Virtual Reality, Archeology and Cultural Heritage (VAST 2005)*, 1–8.

DE HERAS CIECHOMSKI, P., SCHERTENLEIB, S., MAÏM, J., AND THALMANN, D. 2005. Reviving the roman odeon of aphrodisias: Dynamic animation and variety control of crowds in virtual heritage. In *VSMM: Virtual Systems and Multimedia*, 601–610.

DUDASH, B. 2007. Skinned instancing. In *NVidia white paper*.

FARAH, M. J. 1992. Is an object an object an object? Cognitive and neuropsychological investigations of domain specificity in visual object recognition. *Current Directions in Psychological Science 1*, 5, 164–169.

GOSSELIN, D., SANDER, P., AND MITCHELL, J. 2005. Drawing a crowd. *ShaderX3*, 505–517.

HENDERSON, J. M., AND HOLLINGWORTH, A. 1998. Eye movements during scene viewing: An overview. *Eye Guidance in Reading and Scene Perception*, 269–293.

HENDERSON, J. M. 1992. Object identification in context: the visual processing of natural scenes. *Canadian Journal of Psychology 46*, 3, 319–41.

HOCHSTEIN, S., BARLASOV, A., HERSHLER, O., AND SHNEOR, A. N. S. 2004. Rapid vision is holistic. *Journal of Vision 4*, 8, 95.

HOWLETT, S., HAMILL, J., AND O'SULLIVAN, C. 2005. Predicting and evaluating saliency for simplified polygonal models. *ACM Transactions on Applied Perception 2*, 3, 286–308.

ITTI, L., KOCH, C., AND NIEBUR, E. 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence 20*, 11, 1254–1259.

JOHNSON, K. L., AND TASSINARY, L. G. 2005. Perceiving sex directly and indirectly: Meaning in motion and morphology. *Psychological Science 16*, 11, 890–897.

LAARNI, J., RAVAJA, N., AND T. SAARI, T. 2003. Using eye tracking and psychophysiological methods to study spatial presence. In *Proceedings of the 6th International Workshop on Presence*.

LEE, C. H., VARSHNEY, A., AND JACOBS, D. 2005. Mesh saliency. *ACM Transactions on Graphics 24*, 3, 659 – 666.

MAÏM, J., HAEGLER, S., YERSIN, B., MUELLER, P., THALMANN, D., AND VANGOOL, L. 2007. Populating ancient pompeii with crowds of virtual romans. In *Proceedings of the 8th International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST 2007)*, 26–30.

MCDONNELL, R., LARKIN, M., DOBBYN, S., COLLINS, S., AND O'SULLIVAN, C. 2008. Clone attack! Perception of crowd variety. *ACM Transactions on Graphics 27*, 3, 1–8.

NOTHDURFT, H. C. 1993. Faces and facial expressions do not pop out. *Perception 22*, 11, 1287–1298.

NOTON, D., AND STARK, L. W. 1971. Scanpaths in eye movements during pattern perception. *Science 171*, 3968, 308–311.

PETERS, R. J., AND ITTI, L. 2006. Computational mechanisms for gaze direction in interactive visual environments. In *ETRA '06: Proceedings of the 2006 symposium on Eye tracking research & applications*, 27–32.

SINHA, P., BALAS, B., OSTROVSKY, Y., AND RUSSELL, R. 2006. Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE 94*, 11, 1948–1962.

TECCHIA, F., LOSCOS, C., AND CHRYSANTHOU, Y. 2002. Visualizing crowds in real-time. *Computer Graphics Forum (Eurographics 2002) 21*, 4, 753–765.

THALMANN, D., AND MUSSE, S. R. 2007. *Crowd Simulation*. Springer-Verlag.

THALMANN, D., O'SULLIVAN, C., YERSIN, B., MAÏM, J., AND MCDONNELL, R. 2007. Populating virtual environments with crowds. *Eurographics Tutorials 1*, 25–124.

YARBUS, A. L. 1967. *Eye Movements and Vision*. Plenum Press, New York.

YEE, H., PATTANAIK, S., AND GREENBERG, D. P. 2001. Spatiotemporal sensitivity and visual attention for efficient rendering of dynamic environments. *ACM Transactions on Graphics 20*, 1, 39–65.