Variety Is the Spice of (Virtual) Life

Carol O'Sullivan

GV2: Graphics, Vision and Visualisation Group
Trinity College Dublin

Abstract. Before an environment can be populated with characters, a set of models must first be acquired and prepared. Sometimes it may be possible for artists to create each virtual character individually - for example, if only a small number of individuals are needed, or there are many artists available to create a larger population of characters. However, for most applications that need large and heterogeneous groups or crowds, more automatic methods of generating large numbers of humans, animals or other characters are needed. Fortunately, depending on the context, it is not the case that all types of variety are equally important. Sometimes quite simple methods for creating variations, which do not over-burden the computing resources available, can be as effective as, and perceptually equivalent to, far more resource-intensive approaches. In this paper, we present some recent research and development efforts that aim to create and evaluate variety for characters, in their bodies, faces, movements, behaviours and sounds.

1 Introduction

Virtual characters can now be rendered and animated with increasing levels of realism for movies and games. However, no matter how compelling these characters appear individually, once two or more are simulated to form groups or crowds, new challenges emerge. These challenges may be in the way they physically interact with or react socially towards each other, but a particular challenge is to ensure that each individual character looks unique, i.e., that it does not appear to be obviously replicated or *cloned* many times in a scene. There is little more disturbing for the viewer than to see multiple instances of the same character performing the same actions and repeating the same noises or phrases over and over again. In the movie Madagascar, for example, artists created five different kinds of lemurs with 12 variations of hair type, giving 60 unique combinations [13] to create the illusion of a large group of individual animals. However, if the goal is to create realistic humans, creating a crowd of unique and realistic characters is extremely expensive in terms of people hours, memory resources and run-time computation costs.

Therefore, short of manually creating and/or capturing new models, motions and sounds for every single character, some level of cloning is required in most practical applications. Fortunately, tricks can be used to disguise cloned characters, perhaps exploiting graphics hardware to support the rapid generation of
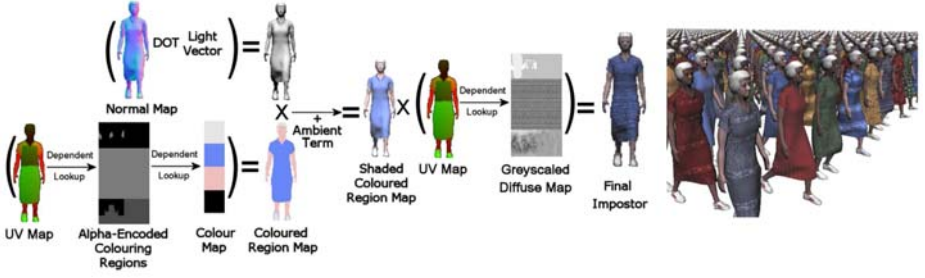
Fig. 1. (left) A typical rendering pipeline that supports variation, and (right) an example of adding variation to one character using 8 different diffuse textures [9][10]

certain types of character variations [14]. For example, the types of variations that are important at the *macroscopic* level i.e., when viewing a crowd of people at a distance, are likely to be much different from *microscopic* factors, i.e., when viewing characters up close. As in most graphical applications, perceptual considerations are very important, and in our recent work we investigated what the effect of different variety schemes are, and how best to balance the need for variety with the practical considerations of the application [22][23].

1.1   Colours and Textures

The easiest and least resource intensive measure for hiding cloned characters in a crowd is simply to vary their colours and/or textures. Using perceptual studies where people were asked to find clones in a scene as quickly as possible, we have found that both types of variation are effective at differentiating human character models who are otherwise identical. In other words, people found such 'colour clones' harder to find in a scene [22], while varying the colours and textures of only the top garments of characters was found to be as effective as full variation [23] (see Figure 3). Hardware accelerated per body part colour modulation [8][9][30] and texture variation [7][10] are two of the most popular methods for variation in real-time crowd applications. Figure 1 shows a typical pipeline, while Figure 2 shows a typical crowd (from our Metropolis project) rendered in realtime using similar techniques.

1.2   Bodies

Varying body shapes is another means of generating lots of novel details in crowds. For example, a number of template human body scans could be used to build a space of human body shapes, where parameterisation could allow novel models to be created by morphing between these templates [1]. Several more such approaches have been proposed (e.g., [2][3][28]). However, such acquired body data has not yet been applied successfully in crowd simulation, perhaps due to the difficulties with rendering, clothing and animating the models effectively.

Fig. 2. A crowd scene from our *Metropolis* crowd system, where colour and texture variation have been used to generate a heterogeneous crowd

Nevertheless, with recent advances (e.g., [16]), the application to crowd simulation is rapidly becoming feasible. It still remains to be seen, however, whether the resulting body shapes and their retargetted motions are actually realistic enough for applications such as games and movies. Another useful source of variation is in the sex [21] (see Figure 4), age or emotional posture of a character (e.g., by making them more stooped or straight).

1.3   Faces

In our recent eye-tracking study of saliency in crowd scenes, we found that people look almost exclusively at the faces (and upper torsos) of virtual humans, when asked to determine whether clones are present in a scene or not [23]. In a set of validation experiments, we found that varying the textures of faces, e.g, adding beards and make-up (see Figures 3, 5 and 6), was as effective as more computationally expensive methods that varied the geometry of the face. These results are consistent with those reported by Sinha et al. [29], who present an overview of important features for face recognition. However, they explain that the shape of the head is also an important cue for recognition. There is clearly scope for further investigation of these effects and their potential exploitation for procedural facial variation.

Also important here is the effect of macroscopic and microscopic factors. At what distance is variation of external facial features (e.g., hair, head shape,
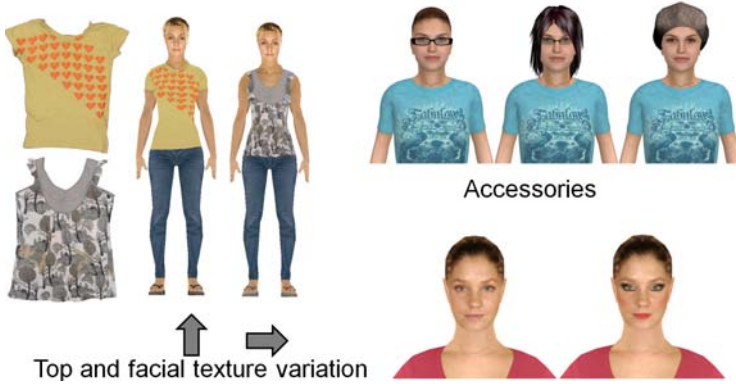
Fig. 3. Disguising clones using texture variations of top garments and faces is cheap and effective. Accessories, though more expensive, also add significantly to the perceived variety of a crowd [23].

Fig. 4. Creating variety by varying the 'male-ness' and 'female-ness' of virtual characters [21]

beard, and jaw-line) needed to make the identities of crowd members look different, and when are similarities between more subtle internal facial features more noticeable (e.g., noses, eyebrows, eyes, mouth)? In previous research designed to assess how the relative contributions of internal and external features change as a function of image resolution, it was found that the two feature sets reverse in importance as resolution decreases [19]. At what distance does facial variation cease to be of any importance and other cues become more informative?

1.4   Animations

So far we have only discussed varying the physical appearance of virtual characters - however, how they *move* is also important, as a population of people all walking in the same way will not be very compelling. Nevertheless, in our studies we have found that appearance clones are much easier to detect than motion clones [22], and even something as simple as making the characters walk out of step can make the task of finding clones significantly harder. This suggests

Fig. 5. Possible facial variations using hair, facial texture and accessories

that more resources should be expended on changing physical appearance than motions and it may be the case that, at the macroscopic level, using only a small number of clips (e.g., one female and one male) could be sufficient. However, as the number of motion clones and the time spent viewing a scene increases, so too do the detection rates. Therefore, methods for varying the motions of individuals are also desirable.

The most natural movements for virtual characters are derived from motion capture systems, which produce a large amount of data. Therefore, optimising the balance between realism, storage and character reactivity is an important issue. Finding a way to create a variety of realistic motions from a smaller subset of real motions is particularly challenging, and it is not always the case that the resulting motions are realistic. Furthermore, certain anomalies can be particularly noticeable, e.g., when a small and thin person's walk is applied to a model of somebody taller and stockier, or when a man's walk is applied to a female model [21].

There has been research in the animation and psychology communities into creating morphed motions for human characters. Giese and Poggio represent complex motion by linear combinations of a small number of prototypical image

Fig. 6. A crowd with variations in colour, textures (top garments and faces) and accessories (hair, hats and glasses)

sequences (e.g., walking, running or limping motion combinations) [15]. Brand and Hertzmann's style machines consist of a learned model that can be used to synthesise novel motion data that interpolate or extrapolate styles, such as ballet or modern dance [5]. Hsu et al also used learning to translate the style of one motion to another, while preserving the overall properties of the original motion - e.g., they could transform a normal walk into one with a sneaky crouch [18]. However, all of these approaches involve interpolation between quite different types or styles of motion, and do not provide a full solution to the more difficult problem of creating novel motions that have the subtle characteristics that make each person individual.

Furthermore, the perceptual responses of the viewer should be taken into account, for example: to what extent can variety be added to a motion without changing its meaning? What are the factors that determine whether a motion appears realistic or consistent for a character? What is the relationship between physical correctness of motion and our perception of that motion? When composing motions from sub-parts (from potentially different characters) what are the rules for deciding whether the composition is realistic or plausible [17]? Such questions pose interesting challenges for the future.

1.5   Behaviours

Ulicny and Thalmann introduced the idea of *levels of variety* [32], where they define the lowest level (LV0) as providing only a single solution for each action; the next level (LV1) allows a choice from a finite number of solutions; while the highest variety level (LV2) allows an infinite number of choices. They note the challenge of applying this concept to traditional AI approaches in crowd

Fig. 7. Varying conversational behaviours to enrich populated scenes

simulation, such as path planning, where the goal is to find an optimal path for an agent to get from one point to another. Without introducing variety, a crowd of virtual humans could end up queuing to get to the end-point, rather than behaving in a more natural way. They overcome this problem by specifying way-points for agents that are randomly distributed within a region close to the path nodes, thus introducing some variety in the paths followed. Another possible approach is to take individual preferences and knowledge of the environment into account when choosing a path for an agent [26]. Recently, Paris and Donikian presented a cognitive architecture to simulate behaviours for fully configurable autonomous agents, thus allowing variety in the decision-making process for high level goals as well as reactive behaviours [25]. Variety can also be added to crowd simulations by assigning individual characteristics and personality traits to agents (e.g., altruistic, dependent, rude, tolerant) [6][11].

As before, investigating how humans perceive crowds can help us to determine how plausible or believable different behavioural simulations (both at a macroscopic and microscopic level) are. We have investigated some of the factors that affect the perception of pedestrian formations in different contexts [27][12]. We have also examined how to add variety in conversational groups and the effects on the perception of realism when individuals' motion clips from several conversations are 'mixed and matched' to generate a variety of different conversational groups [20] (see Figure 7).

1.6   Sounds

Finally, for a crowd to be realistic they should be making appropriate crowd sounds. Recent approaches to audio rendering of complex scenes use perceptual metrics to decide how to cluster multiple sound sources and how to selectively allocate resources (e.g., to visible sources) [24][31]. As with models and

motions, capturing, storing and rendering sounds presents a significant burden on any system. Representing, clustering and synthesizing new speaking sounds is acknowledged to be an especially difficult case. Statistical parametric speech synthesis, especially approaches based on Hidden Markov Models (HMM) have recently been used to create a variety of different voices [4]. It remains to be seen if such methods could be used to create a large variety of voices in a crowd, or if indeed this level of variety is important for speech, depending on the task. At what point does an individual voice become important? Are people more or less sensitive to variations in voices than in appearance or motions, and how do these factors interact? (So far, we have found in our studies of appearance and motion that the former dominates - will sound affect this?)

Finally, synchronising congruent sounds with character animations is particularly challenging. For example, at a microscopic level footsteps should be synchronised with the footplants of an animated character, and the sound should be congruent with their appearance - e.g., a heavy person with boots should sound different from a woman with high heels, or an appropriate conversational sound should emanate from a person talking as they pass the camera. Should speech also be fully synchronised with the facial animation and gestures of each up-close character? When are such details no longer necessary, as at a macroscopic level e.g., for large crowds viewed from a distance, they almost certainly would not be discernable.

1.7   Conclusions

In this paper, an overview has been presented of current research into creating variety for virtual humans, with a particular emphasis on crowd simulation applications. It is certainly not intended to be an exhaustive survey of the field, but rather to highlight some of the challenges that face researchers and practitioners. While advances have been made in each individual domain of modelling, rendering, animation, behaviour simulation and audio synthesis, further interesting research directions will involve combining and synchronising the variations created to deliver a coherent and rich multisensory percept.

Acknowledgements

References

1. Allen, B., Curless, B., Popović, Z.: The space of human body shapes: reconstruction and parameterization from range scans. In: ACM SIGGRAPH 2003, pp. 587–594 (2003)

2. Allen, B., Curless, B., Popović, Z., Hertzmann, A.: Learning a correlated model of identity and pose-dependent body shape variation for real-time synthesis. In: Proceedings of the 2006 ACM SIGGRAPH/Eurographics symposium on Computer animation, pp. 147–156 (2006)
3. Anguelov, D., Srinivasan, P., Koller, D., Thrun, S., Rodgers, J., Davis, J.: Scape: shape completion and animation of people. ACM Transactions on Graphics 24(3), 408–416 (2005)
4. Black, A.W., Zen, H., Tokuda, K.: Statistical parametric speech synthesis. In: Acoustics, Speech and Signal Processing, 2007, vol. 4, pp. IV–1229–IV–1232 (2007)
5. Brand, M., Hertzmann, A.: Style machines. In: ACM SIGGRAPH 2000, pp. 183–192 (2000)
6. Braun, A., Musse, S.R., de Oliveira, L.P.L., Bodmann, B.E.J.: Modeling individual behaviors in crowd simulation. In: Computer Animation and Social Agents (CASA 2003), p. 143. IEEE Computer Society, Los Alamitos (2003)
7. De Heras, C.P., Schertenleib, S., Mäim, J., Thalmann, D.: 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), pp. 1–8 (2005)
8. De Heras, C.P., Schertenleib, S., Mäim, J., Thalmann, D.: Reviving the roman odeon of aphrodisias: Dynamic animation and variety control of crowds in virtual heritage. In: VSMM: Virtual Systems and Multimedia, pp. 601–610 (2005)
9. Dobbyn, S., Hamill, J., O'Conor, K., O'Sullivan, C.: Geopostors: a real-time geometry / impostor crowd rendering system. ACM Transactions on Graphics 24(3), 933 (2005)
10. Dobbyn, S., McDonnell, R., Kavan, L., Collins, S., O'Sullivan, C.: Clothing the masses: Real-time clothed crowds with variation. In: Eurographics Short Papers, pp. 103–106 (2006)
11. Durupinar, F., Allbeck, J., Pelechano, N., Badler, N.: Creating crowd variation with the ocean personality model. In: AAMAS 2008: Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems, pp. 1217–1220 (2008)
12. Ennis, C., Peters, C., O'Sullivan, C.: Perceptual evaluation of position and orientation context rules for pedestrian formations. In: APGV 2008: Proceedings of the 5th symposium on Applied perception in graphics and visualization, pp. 75–82 (2008)
13. Galli, P.: Madagascar tech turns imagination into reality (2005), eWeek.com
14. Galvao, R., Laycock, R., Day, A.M.: Gpu techniques for creating visually diverse crowds in real-time. In: VRST 2008: Proceedings of the 2008 ACM symposium on Virtual reality software and technology, pp. 79–86 (2008)
15. Giese, M., Poggio, T.: Morphable models for the analysis and synthesis of complex motion patterns. International Journal of Computer Vision 38(1), 59–73 (2000)
16. Hasler, N., Stoll, C., Sunkel, M., Rosenhahn, B., Seidel, H.-P.: A statistical model of human pose and body shape. Computer Graphics Forum (Eurographics 2009) 28(2), 337–346 (2009)
17. Heck, R., Kovar, L., Gleicher, M.: Splicing upper-body actions with locomotion. Computer Graphics Forum (Eurographics 2006) 25(3), 459–466 (2006)
18. Hsu, E., Pulli, K., Popović, J.: Style translation for human motion. ACM Transactions on Graphics (SIGGRAPH 2005) 24(3), 1082–1089 (2005)
19. Jarudi, I.N., Sinha, P.: Relative contributions of internal and external features to face recognition. Technical Report CBCL Paper #225/AI Memo #2003-004, Massachusetts Institute of Technology (2003)

20. McDonnell, R., Ennis, C., Dobbyn, S., O'Sullivan, C.: Talking bodies: Sensitivity to de-synchronisation of conversations, vol. 6(4) (to appear, 2009)
21. McDonnell, R., Jörg, S., Hodgins, J.K., Newell, F., O'Sullivan, C.: Evaluating the effect of motion and body shape on the perceived sex of virtual characters. TAP 5(4), 1–14 (2009)
22. McDonnell, R., Larkin, M., Dobbyn, S., Collins, S., O'Sullivan, C.: Clone attack! perception of crowd variety. ACM Transactions on Graphics (SIGGRAPH 2008) 27(3) (2008)
23. McDonnell, R., Larkin, M., Hernández, B., Rudomín, I., O'Sullivan, C.: Eye-catching crowds: saliency based selective variation. ACM Transactions on Graphics (SIGGRAPH 2009) 28(3) (2009)
24. Moeck, T., Bonneel, N., Tsingos, N., Drettakis, G., Viaud-Delmon, I., Alloza, D.: Progressive perceptual audio rendering of complex scenes. In: I3D 2007: Proceedings of the 2007 symposium on Interactive 3D graphics and games, pp. 189–196 (2007)
25. Paris, S., Donikian, S.: Activity-driven populace: a cognitive approach for crowd simulation. IEEE Computer Graphics and Applications (Virtual Populace special issue) 29(4), 34–43 (2009)
26. Paris, S., Donikian, S., Bonvalet, N.: Environmental abstraction and path planning techniques for realistic crowd simulation. Computer Animation and Virtual Worlds 17, 325–335 (2006)
27. Peters, C., Ennis, C.: Modelling groups of plausible virtual pedestrians. IEEE Computer Graphics and Applications (Virtual Populace special issue) 29(4), 54–63 (2009)
28. Seo, H., Cordier, F., Magnenat-Thalmann, N.: Synthesizing animatable body models with parameterized shape modifications. In: ACM SIGGRAPH/Eurographics Symposium on Computer Animation, pp. 120–125 (2003)
29. Sinha, P., Balas, B., Ostrovsky, Y., Russell, R.: Face recognition by humans: 19 results all computer vision researchers should know about. Proceedings of the IEEE 94(11), 1948–1962 (2006)
30. Tecchia, F., Loscos, C., Chrysanthou, Y.: Visualizing crowds in real-time. Computer Graphics Forum (Eurographics 2002) 21(4), 753–765 (2002)
31. Tsingos, N., Gallo, E., Drettakis, G.: Perceptual audio rendering of complex virtual environments. In: ACM SIGGRAPH 2004, pp. 249–258 (2004)
32. Ulicny, B., Thalmann, D.: Towards interactive real-time crowd behavior simulation. Computer Graphics Forum 21(4), 767–775 (2002)